

## Timely messages for the South

**A WHO report delivers a powerful combination of analysis and economic judgement to suggest that developed countries have much to gain by significantly boosting their efforts to alleviate health problems in the developing world.**

**Q**uadruple the developed world's donations to the developing world to improve health. Double the aid by governments, charities and industry money devoted to R&D on diseases that particularly affect poor countries. Increase by whole percentages of GNP the funds spent by developing countries on their own peoples' health. And sustain these levels of support for the next two decades.

These are no small recommendations. They come from the hope-fully influential Commission on Macroeconomics and Health, set up two years ago by the head of the World Health Organization, Gro Harlem Brundtland. What is depressing is how implausibly challenging those goals seem — and yet how essential, if combating the disease burdens of the great majority of the world is to be seriously addressed.

We have been here before. As long ago as 1980, the Brandt Commission on International Development Issues called for much greater support from “the North” for (among other things) health in “the South” by 2000. Over a decade ago, the Global Health Forum highlighted the fact, still uncomfortably true, that only about 5% of what is now \$60 billion per year spent globally on biomedical research is directed at diseases affecting the poorest 95% of the world's population. How can one persuade a reluctant world to accept yet another set of recommendations addressing that huge imbalance?

One of the new report's main strengths (see [www.who.int](http://www.who.int)) is its attention to the reality of healthcare delivery. It identifies the need for enhanced research into operational aspects of healthcare, as well as the more traditional areas of epidemiology and basic biomedical research. It highlights the pitifully small degree of success by those in developed countries in raising funds for research in major diseases that kill millions every year — even malaria and tuberculosis have attracted a fraction of what is needed for research into drugs and vaccines.

The challenges are not only financial but also organizational, both in healthcare delivery and research. Indeed, one of the report's messages for agencies funding biomedical research is relevant to issues closer to home — the need to coordinate expertise internationally and to capture data in a way that ignores national boundaries, as well as the need to help young researchers. With these goals in mind, national biomedical fund-

ing agencies need to standardize their application procedures and peer-review processes, and increase the entitlement of foreigners to apply for funds. But the commission also recommends setting up a Global Health Research Fund with an annual budget of \$1.5 billion. Progress already made by a similar organization, the Consultative Group on International Agricultural Research, adds weight to this proposal.

The commission displays excessive casualness towards one aspect of research. It gives due attention to the need to mobilize funds while allowing pharmaceutical companies to protect their interests through the patent system. In contrast, it also urges the free instantaneous distribution of scientific information without acknowledging the demonstrable obstacles to such a course of action — an area where some fresh economic insights could have been helpful. Nevertheless, the underlying message needs to be heeded: publishers, like pharmaceutical companies, can do more than they already are to enhance access to the literature by researchers in the developing world, through imaginative partnerships and targeted pricing.

The commission's analysis highlights the possibilities for both developed and developing countries to do more. It also identifies the economic incentives for them to do so. Its recommendations if implemented would, it says, save eight million lives per year by 2010, or 330 million “disability-adjusted life years”, which would directly release at least \$180 billion of earnings, plus a move away from the poverty trap and a substantial consequent growth in economies and international market opportunities.

Researchers and policy-makers can do more than some might imagine. One thinks of scientists who recruited the downtime of people's desktop computers to hunt for extraterrestrial intelligence or, more relevant, potential targets for drugs; or policy-makers who conceived programmes to foster scientific collaborations between politically divided nations. But there is also a message that is both moral and self-interested for developed countries. To quote the report: “Governments and agencies who help stimulate and nurture these actions will be providing a specific antidote to the despair and hatred that poverty can breed.” Who could say anything but “Amen” to that? ■

## Towards more effective drug discovery

**That's a key goal of a new Nature journal.**

**T**here is little doubt that the drug-discovery business faces trouble ahead. Despite huge increases in research spending, the pharmaceutical industry produces roughly the same number of new drugs each year, and there is little evidence that things are about to get better. Arguably, a lack of communication is one reason for this failing. In all the excitement over the new technologies which it was hoped would revolutionize drug discovery, people seem to have forgotten that the discovery process is an integrated business. Although thousands of novel targets will be revealed over the coming years, a lead compound with affinity for an isolated protein is a very long way from a drug. Looking back over case histories reveals that most successful drug discoveries depended on a great deal of

cross-fertilization between disciplines — chemists, biologists, toxicologists and clinicians all need to be speaking the same language.

Launched this month, *Nature Reviews Drug Discovery* aims to capture the growing realization among members of the drug-discovery community that they need to start talking to each other again and, in the process, to appeal to both academic and industrial audiences. Editorially independent of *Nature* (as are all *Nature* journals), it is described by its editors as a “guide to the evolving world of drug discovery”. If, in pursuit of that goal, it cuts through the hype surrounding many new areas of development, and gives a clear picture of what drug discovery can be at its best, it should indeed provide a valuable service. ■





**Web surfing**  
Planned database will compile research on marine mammals  
**p4**



**Tribal dispute**  
Anthropologists disagree over how to respond to criticism  
**p5**



**Not in my cave**  
Nevada struggles to block huge nuclear waste dump  
**p6**



**In the clear**  
Copyright charges against computer scientist are dropped  
**p6**

# German task force outraged by changes to science fraud report

**Quirin Schiermeier, Munich**

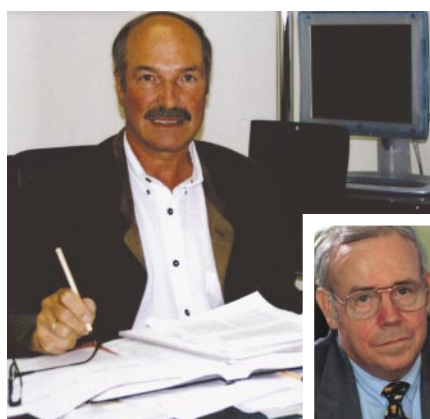
Investigations into Germany's largest case of scientific misconduct have ended in shrill discord. Members of an independent task force set up to investigate the scandal by the DFG, Germany's main research granting body, are now complaining that the agency has watered down their report.

Accusations of scientific fraud were first raised in 1997 against two cancer researchers, Friedhelm Herrmann and Marion Brach, who worked together at the Max Delbrück Centre for Molecular Medicine in Berlin in the early 1990s. As investigations continued, allegations of data fabrication extended to more and more papers. Together with subsequent scandals, the case has sparked fears that misconduct may be rife in German clinical research.

In its final report into the Herrmann and Brach affair, published in June 2000, the six-member task force concluded that data in at least 94 papers had either definitely or "highly probably" been manipulated (see *Nature* **405**, 871–872; 2000). But since then, a bitter row has been simmering between the task force's members and the DFG.

Under pressure from the lawyers of one researcher under suspicion, Lothar Kanz, now at the University of Tübingen, the DFG deleted passages from the task force's report describing alleged irregularities in two papers, published in *Blood* (**84**, 1421–1426; 1994) and *The New England Journal of Medicine* (**333**, 283–287; 1995). The papers were co-authored by Kanz, his Tübingen colleague Wolfram Brugger and Roland Mertelsmann of the University of Freiburg.

The dispute has boiled over in recent weeks. Ulf Rapp, a cell biologist from the University of Würzburg and head of the task force, complained about the "unauthorized manipulation" of the task force's report in a 19 November letter, signed by four other task-force members, to DFG president Ernst-Ludwig Winnacker. In his response, Winnacker defended the deletions, arguing that the press conference that accompanied the report's publication would otherwise have had to be cancelled. "This would have been regarded in public as a victory for the



**Taken to task:** Rapp has complained to Winnacker (inset) about changes to the task force's findings.

scientists accused of scientific misconduct," he claimed.

Rapp replied in a letter on 17 December: "By playing down striking manipulations of data, the DFG has departed altogether from its role as keeper of good scientific practice." He now says that "the matter is likely to be finished up in court", and argues that the DFG did not have the authority to alter the task force's report.

The DFG eventually banned Kanz and

Brugger from serving as peer reviewers for two years, and has issued new guidelines on good scientific practice and clinical research. Mertelsmann, a prominent gene therapist, received a five-year ban.

Rapp feels disillusioned by his experience. "I now understand that the DFG represents the state of affairs in clinical research, and that you cannot expect much more from it than lip service," he told *Nature*. "There must be more transparent ways of making public what goes wrong. Sweeping things under the carpet because of fear of being sued is not an option."

Rapp wants the DFG to publish an index, in English, of all contaminated publications, either in an international scientific journal or on the Internet. He also argues that the agency is not the right body to deal with allegations of misconduct. "What we need is a truly independent panel to handle investigations," he says.

Eberhard Hildt, who worked as a postdoc with Herrmann and Brach at the University of Ulm and first raised the alarm over the suspect data, agrees. "It is a shame that papers involving fabricated data are still being quoted, and that young researchers might build on them," he says. ■

## Body identified as missing biologist

**Erika Check, Washington**

**The body of Don Wiley, the prizewinning structural biologist who went missing in November, was found on 21 December. Hydroelectric workers spotted his body snagged on a tree in the Mississippi River, more than 450 kilometres downstream of the bridge near Memphis where his hire car was abandoned (see *Nature* **414**, 475; 2001).**

Wiley, a professor at Harvard University, had won a Lasker award and the Japan Prize for his work on the structure of major histocompatibility complex proteins, which are involved in immune responses. "This is devastating for his family and only slightly less so for our department," says Harvard

colleague Jack Strominger, who shared those awards.

Wiley had last been seen on 15 November at a dinner for St Jude Children's Research Hospital in Memphis, where he served as a member of its scientific advisory board. Those who were present are stunned by Wiley's death. He was "in great spirits, talking about science, art, politics and his family", says William Evans, deputy director of the hospital. "To us who were with him and who knew him, we felt it was inconceivable that he could do any harm to himself."

As *Nature* went to press, no autopsy information had been released. ■



Seal the deal: the new project aims to persuade animal biologists to contribute their tagging data.

## Online marine resource could soon be swimming with data

**Jonathan Knight, Santa Cruz**

Decades of raw data on the movements and diving behaviour of dolphins, whales, seals and other marine animals could soon be made available in an online database.

Leading figures in the field met last month to start laying plans for this resource. "Basically, we are trying to make a GenBank for diving data on marine mammals," says Dan Costa, a marine biologist at the University of California, Santa Cruz.

Researchers have been using time-depth recorders (TDRs) to track the movement of ocean mammals since the 1960s. TDRs adhere to an animal's back for days or weeks and then float to the surface for recovery. More recently, satellite transmitters that provide information on latitude and longitude have also become available. As they have become smaller and cheaper, the number of researchers using them has mushroomed.

Recently, tags have shown where blue whales feed and calve (B. A. Lagerquist, K. M. Stafford & B. R. Mate *Mar. Mamm. Sci.* **16**, 375–391; 2000) and revealed the remarkable diving ability of seals and sea lions (D. P. Costa, N. J. Gales & M. E. Goebel *Comp. Biochem. Physiol. A* **129**, 771–783; 2001).

Some of the data are already stored in public databases, but the vast majority are analysed once and then shelved. Placing them in a publicly accessible database would help biologists who want to develop a more complete picture of ocean ecosystems.

But several hurdles remain, not least convincing marine biologists to release their data. Standards for organizing data will also be needed to allow searching and comparison of different tracking and dive records.

To make a start on these issues, Costa and

his postdoc Scott Shaffer gathered 45 marine biologists in Santa Cruz. Participants agreed on the utility of a data archive, but few are likely to contribute data they have not yet published.

"Investigators want to know they will have first dibs," says Bruce Mate, a whale biologist at the Hatfield Marine Science Center of Oregon State University in Newport.

So, rather than collecting tracking data as it is generated, the archive will, at least initially, be restricted to data from published studies. The resulting lag time of a year or two from data acquisition to its appearance in the archive will not be a serious problem, Mate argues, because the database's main value will be in providing access to information that is currently unavailable. Published papers contain only data summaries.

Data comparison will also be a challenge. There are several manufacturers of TDRs, and their products record essentially the same information but in different formats. Researchers may also calibrate instruments differently, setting them up to record only longer dives, for example, something that would have to be indicated in the database.

Although the archive will deal initially with marine mammals, the goal is to expand it to include other groups. Fish and seabird researchers are keen to join in, says Barbara Block, also at the University of California, Santa Cruz, who tags a range of marine vertebrates including tuna, albatrosses and turtles.

Costa's first step will be to set up a website detailing what data have been collected and where they are held. He also plans to develop a formal proposal for funding, possibly for submission to the US Office of Naval Research, which backed last month's meeting.

## Future funds in doubt as asteroid project wins short reprieve

**Erika Check, Washington**

Following an outcry from planetary scientists, NASA has abandoned a plan to terminate radar studies of near-Earth asteroids and planets using the world's largest radio telescope. But the project's future remains uncertain, as it has only been guaranteed one more year of funding.

On 10 December, NASA told scientists working on the studies using the Arecibo Observatory in Puerto Rico that they would lose their \$550,000 funding in 2002. But just 10 days later, the agency provided \$400,000 after protests from the Planetary Society and researchers in the field.

The radar studies are supported from the same budget that finances the search for new near-Earth asteroids. But NASA is concerned that the annual total of \$3.55 million cannot cover both the discovery of new asteroids and studies of old ones.

Don Campbell, head of Arecibo's Radar Astronomy Group, claims the telescope is crucial to understanding near-Earth objects. Its radar equipment measures their shape, surface properties and velocity. "This information can be used to make substantial improvements in our ability to predict the future orbits of these objects," he says. The only comparable facility is at Goldstone, in California's Mojave Desert, but its equipment is less sensitive than that at Arecibo.

However, Tom Morgan, discipline scientist for NASA's Near-Earth Object Program, argues that space missions will prove to be "the best way to go out and answer the questions we need to answer if, God forbid, we ever have to change the orbit of one of these objects."

From 2003, the radar project will have to pass a round of peer review if it is to continue receiving funds. NASA may also try to shift the studies to the National Science Foundation, which provides most of Arecibo's \$11-million budget.

www.naic.edu



Scope for concern: NASA may struggle to fund continued study of asteroids at Arecibo.



# Anthropologists split over misconduct claims

Rex Dalton, San Diego

One year after a journalist rocked the world of anthropology with allegations of serious misconduct in research among Amazonian tribes, the American Anthropological Association (AAA) remains deeply divided over how to respond. In the latest twist, the society has been forced to remove a draft report on the allegations from its website, following dissent among contributors over its contents.

In his book *Darkness in El Dorado: How Scientists and Journalists Devastated the Amazon*, Patrick Tierney examined three decades

of research among the Yanomami people in Venezuela. He alleged that geneticist James Neel of the University of Michigan and anthropologist Napoleon Chagnon of the University of California, Santa Barbara, improperly bartered weapons for interviews, conducted clinical studies without informed consent, and may have exposed the Yanomami to infectious diseases (see *Nature* 408, 391; 2000). Neel died in 2000 before the book was published, but Chagnon, who is retired, has protested his innocence.

The AAA subsequently set up a task force

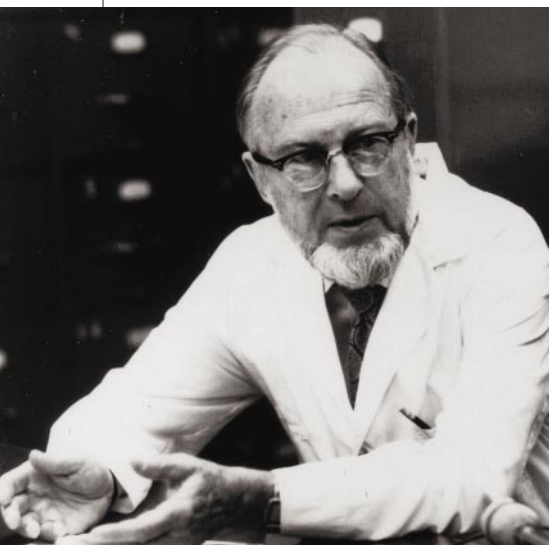
to investigate the charges and to draw up new standards for the conduct of studies of indigenous peoples. An early draft appeared on the society's website shortly before the AAA's annual meeting in Washington in November. But at that gathering, the society's leaders were heavily criticized over the draft's contents, and for making it available. It was removed from the website shortly afterwards.

At the meeting, Fernando Coronil of the University of Michigan and Janet Chernela of Florida International University near Miami, both members of the task force, complained that they had not seen the document before it was put online. Both argued that the draft failed to accurately reflect the accused researchers' responsibility for identified problems.

The disagreements show how hard it is for scientific societies to deal with allegations of misconduct, particularly when the accused — like Neel and Chagnon — are prominent figures. Jane Hill of the University of Arizona in Tucson, who chairs the task force, says that there is still much work to do. AAA officials hope that the task force will nonetheless complete its report by the spring.

At the AAA meeting, Coronil won support for a proposal allowing members to write and sign individual sections of the report. This, he hoped, would allow the task force "to make a more powerful statement". But whether it will lead to a report that is acceptable to both the task force and the association's members remains unclear. ■

www.aaanet.org



James Neel (above) is accused of improper research into the Amazonian Yanomami people.



## Museum staff strike to prevent loss of artefacts

Many of the staff at France's leading anthropology museum, the Musée de l'Homme in Paris, are on strike in protest at a proposal to move its ethnology collection to a new museum of primitive art and civilizations.

The staff are angry that they may be deprived of access to the artefacts, which represent peoples from around the world. "Without our collections, we can no longer function as a museum, linking teaching and research," claims Bernadette Robbe, a researcher at the museum's ethnology laboratory, and a member of the strike committee.

The backdrop to the strike is a widely accepted need to reform the museum, which has failed to modernize the presentation of its collections and has slipped back from the forefront of anthropological research. After much debate, a commission set up by French president Jacques Chirac decided to transfer

the ethnology collection, the museum's library and its photographic collections, to a new museum, due to open in 2004. This will be known as the Musée du quai Branly, after its riverside location in the east of Paris.

But Robbe argues that no decisions about the collections should have been taken without consultation with the new president and scientific advisory committee of the National Natural History Museum — the Musée de l'Homme's parent body. As *Nature* went to press, appointments to these posts had not been made, but were expected imminently. Robbe accepts that the museum needs to be reformed, but says that lack of money and personnel have prevented staff from taking their ideas forward.

Jean-Claude Moreno, the interim administrator of the natural history museum, says that the strike is symptomatic of the museum's "intellectual and scientific crisis". He argues that the museum has

missed at least three opportunities to reform its research programme, because staff could not agree on a coherent scientific project.

The decision to move the collections has provoked mixed responses in the wider research community. William Sturtevant, curator of North American ethnology at the Smithsonian Institution's National Museum of Natural History in Washington, describes the collections as among the best half-dozen in the world. "The proposed change is potentially disastrous for preservation of and access to the artefacts," he says.

But Dominique Michelet, deputy director of the unit of Archaeology of the Americas in Paris, part of the CNRS, France's national research agency, says: "This is an unmissable opportunity to restore the collections and recreate a strong research base." ■

www.mnhn.fr/mnhn/mdh

www.quaibranny.fr/00/dynamique/home.html/an

## Big increases in funding for NIH and bioterrorism labs

**Washington** Funding for the US National Institutes of Health (NIH) is to increase by 15%, after Congress decided on 20 December to raise the institutes' budget for this year to just over \$23 billion.

The budget — part of a larger bill on education, labour and health — had been held up by debates over matters unrelated to the NIH. Research leaders and lobbyists had been confident of a substantial increase, but warn that next year will present a bigger challenge. Some fear that the current economic climate may threaten the five-year plan to double the NIH's budget to \$27.3 billion by 2003.

In a separate defence-spending bill, \$170 million was awarded to the National Institute of Allergy and Infectious Diseases in Bethesda, Maryland, for projects to tackle bioterrorism and for the construction of additional biosafety level 4 labs, which are used to study dangerous pathogens. The same bill also allocates \$25 million to improving security at the NIH and the Centers for Disease Control and Prevention (CDC) in Atlanta. The CDC has already been allocated \$100 million to upgrade its research facility in Fort Collins, Colorado.

## Fate of Japan's sleeping satellite is in lap of the gods

**Tokyo** Japan's Yohkoh satellite looks set to sleep through its tenth birthday party. The satellite's observations of the Sun came to a surprise halt on 13 December, just weeks short of this month's planned celebration of a decade of data collecting.

Yohkoh's operators at the Institute of Space and Astronautical Science (ISAS), Japan's main space science organization, were left helpless when Yohkoh's solar panels unexpectedly swung away from the Sun, cutting off the satellite's power supply. The lack of power means that the position of the satellite cannot be corrected.

Engineers now hope that the satellite will reorient itself long enough for its batteries to recharge. "We need to wait for the gods to give us a chance," says Takeo Kosugi, who works on the Yohkoh project. ISAS planned to keep Yohkoh running until 2005, when its successor, Solar-B, is due to be launched.

♦ [www.solar.isas.ac.jp](http://www.solar.isas.ac.jp)

## Nevada takes nuclear waste case to court

**San Francisco** In a last-ditch effort to avoid becoming North America's main repository for nuclear waste, the state of Nevada is suing the US Department of Energy (DOE) over its



**Deep trouble:** plans to store nuclear waste under Yucca Mountain are to be tested in the US courts.

proposed storage facility at Yucca Mountain, 145 kilometres northwest of Las Vegas.

The suit, filed on 17 December, asks the US Court of Appeals in Washington DC to block the project on the grounds that the DOE has changed its own rules. The waste was originally to be sealed off inside caves 300 metres below the mountain. But after studies found that water could seep in from rainfall and dormant volcanoes in the region could erupt, the DOE decided to put the waste inside the caves in metal containers.

According to Nevada state officials, this breaches a requirement of the 1982 Nuclear Waste Policy Act for a permanent repository to rely on geological containment. Energy Secretary Spencer Abraham is to decide whether to recommend the site to President George W. Bush in the coming weeks.

## Standard-model violation 'a mistake', say physicists

**London** Reports of a violation of the standard model of particle physics made last February by a group at the Brookhaven National Laboratory in New York may have their origin in an algebraic error.

The Brookhaven group measured the magnetic moment of a particle known as a muon, and found it to differ from the predicted value. They then claimed that their findings could violate the standard model, although some physicists disputed this (see *Nature* 410, 291; 2001).

However, Marc Knecht and Andreas Nyffeler of the Centre for Theoretical Physics in Marseille have now revealed a mistake in the predictions: theorists were using a value for a factor called the pion pole contribution as  $-55.6 \times 10^{-11}$ , when it should have been  $55.6 \times 10^{-11}$ . Use of the correct value increases the probability that the difference between the theoretical and experimental values could be due to a statistical fluke from 1% to 13%. Nyffeler and Knecht have submitted their findings to the journal *Physical Review D*.

## Bomb tests may have led to pilot's death

**London** British police are investigating claims that a military pilot was ordered to fly his unprotected plane through the mushroom cloud created by a 1958 nuclear-weapons test.

The test was one of several carried out during the 1950s in Australia and small Pacific islands by the British government. Eric Denson, a squadron leader in the Royal Air Force, fell ill in subsequent years and committed suicide almost 20 years after the test. Denson's widow claims that he was illegally exposed to high radiation levels and that this eventually led to his death.

The inquiry — which the police say is a "preliminary assessment" at this stage — could lead to criminal charges against the Ministry of Defence and the scientists involved in the tests. The British government already faces the threat of court action by Australian and New Zealand servicemen who claim to have been exposed to high levels of radiation during the testing programme (see *Nature* 412, 5 & 670; 2001).

## US drops case against Russian programmer

**San Diego** The US government has dropped criminal charges against Dmitry Sklyarov, the Russian computer scientist who was arrested when he attended a Las Vegas hackers' convention in July.

Software written by Sklyarov, known as the Advanced eBook Processor, can be used to read encrypted versions of electronic books. The software was deemed to violate the US Digital Millennium Copyright Act, a controversial law designed to protect digital commerce. Freedom-of-speech advocates argue that the act limits the traditional fair use of copyright materials, such as for education or research purposes.

Charges against Sklyarov were dismissed on 13 December on condition that he agreed to cooperate with the case against the software's distributors Elcomsoft, a Moscow-based computer company where he worked while studying for his doctorate.



**Free speech:** Dmitry Sklyarov, left, with Elcomsoft president Alexander Katalov.



# Surviving a knockout blow

Disabling a gene in one mouse strain can be fatal — but in another strain it can produce animals that seem normal. Making sense of such results requires stamina and skill, says Helen Pearson.

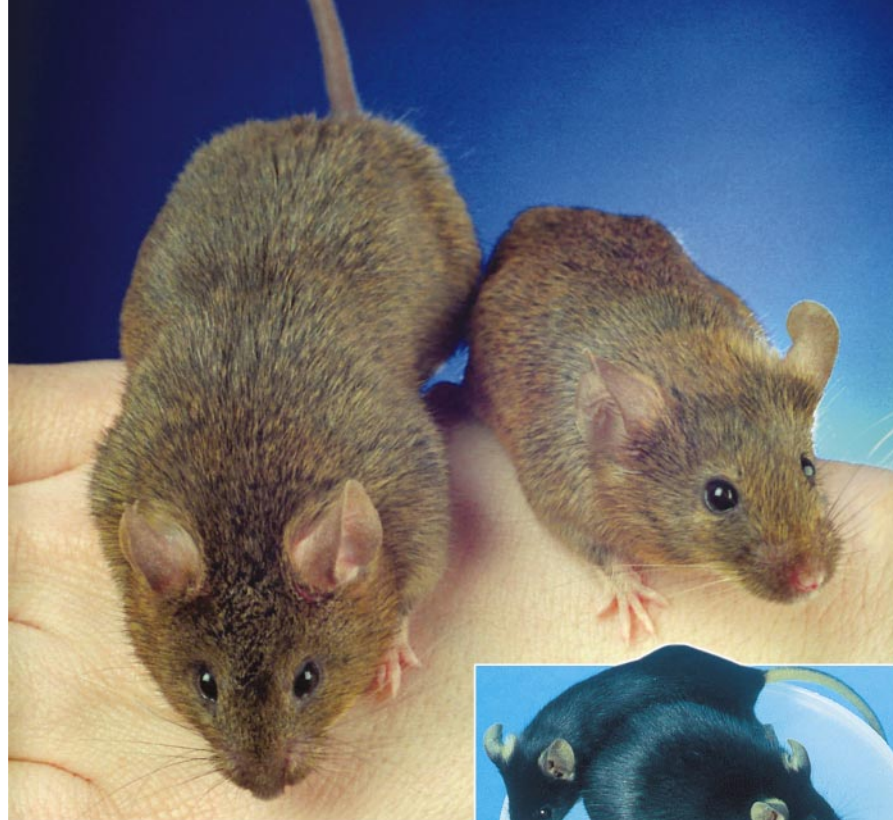
**S**ome mice should, by rights, be dead. At the very least, Teyumuras Kurzchalia expected his to be critically ill. But the most prominent symptom of his genetically engineered mice was a persistent erection.

The mice lacked a gene called *caveolin-1*, which is needed to make the flask-shaped pits that pockmark the surface of many mammalian cells and are thought to help assemble molecules that pass signals to the cellular interior. Disrupting the gene should have caused serious problems, reasoned Kurzchalia, who works at the Max Planck Institute of Molecular Cell Biology and Genetics in Dresden<sup>1</sup>.

Kurzchalia's priapic rodents join a growing menagerie of mutant mice that are causing their makers some bemusement. Since the late 1980s, when 'knockout' technology to selectively disable target genes in mice was developed, it has proved a powerful tool for revealing gene function. "It's still a big dream — that knockouts will explain everything," says Josef Penninger of the Amgen Institute in Toronto, whose lab is one of the most productive in the knockout business.

But from the start, curious results have clouded this vision. In many cases, a mutant mouse does not show any obvious characteristics — or phenotype. In others, the phenotype disappears when the disabled gene is crossed into a different strain of mouse. Indeed, clear and consistent phenotypes now seem to be the exception rather than the rule.

These variable results often reflect the fact that genes acting in parallel pathways can compensate for the one that is missing. In such cases, strain differences reflect the different genetic background in which the disrupted gene finds itself. In other cases,



**Spot the difference:** some mutant mice show clear attributes, but often the genetic effects are hidden.

phenotypes go undetected because the techniques used to look for them are too crude. "We've all been a bit naive," concludes Howard Jacob, who studies the genetics of complex phenotypes at the Medical College of Wisconsin in Milwaukee.

But geneticists argue that with painstaking work, these problems can be overcome — and even be made into a virtue. Variation in results from different strains, for example, provides a starting point from which to track down the compensatory genes.

Better knowledge of the characteristics of common lab strains will also help researchers to decide which mice to conduct their knockout experiments on. And techniques to find subtle phenotypes are being refined — none too soon, because knockout technology is now being joined by huge 'mutagenesis' screens in which chemical mutagens are used to derive thousands of mice in which genes have been altered at random. The first two screens published preliminary results last year<sup>2,3</sup>, and a dozen or so similar projects are in progress.

## Breeding hell

It was Terry Magnuson of the University of North Carolina at Chapel Hill who opened many mouse geneticists' eyes to the influence of the rest of the genome on knockout experiments. In 1995, his team disabled the gene for the epidermal growth-factor receptor. In one mouse strain, CF-1, the knockout embryos perished at around the time of implantation in the uterus. But in the CD-1 strain, they survived for up to three weeks after birth<sup>4</sup>. "From that time on, everyone started paying much more attention to the

genetic background," says Magnuson.

Ideally, experiments on knockout mice would routinely include work on multiple strains. In practice, most researchers in the field argue that this is not realistic — creating a single knockout strain can take up the majority of a three-year PhD project.

But where strain differences are uncovered, there is the potential to use them to track down the 'modifier' genes that enhance or suppress a characteristic. The principle was demonstrated in mice carrying a chemically induced mutation in the tumour-suppressor gene *adenomatous polyposis coli* (*APC*). In the C57BL/6J strain, this mutation causes mice to develop numerous polyps — growths in the colon that can become cancerous. But in the AKR strain, the same mutation caused mice to develop only a handful of polyps. By crossing the two strains together, Bill Dove of the University of Wisconsin in Madison and his colleagues identified marker genetic sequences that were inherited along with the AKR mice's milder phenotype. This narrowed the search for the modifier gene to chromosome four<sup>5</sup>. It was eventually shown by other researchers to encode an enzyme called phospholipase A2, which is involved in the inflammatory response to polyps and is inactive in the C57BL/6J strain<sup>6</sup>.

So far, there have been few successful attempts to follow Dove's lead using knockout mice. The problem is that the experiments are time-consuming and expensive. And if several genes are all influencing the phenotype, each with a relatively small effect, mapping them to a region of the genome becomes extremely difficult<sup>7</sup>.

But interpreting strain variation should

be aided by attempts to collate information on the natural phenotypic differences between commonly used mouse strains. The Mouse Phenome Project, based at the Jackson Laboratory in Bar Harbor, Maine, was launched in 2000 and aims to collect baseline physiological and behavioural data, such as blood glucose levels, growth rate and daily rhythms of activity for 40 different inbred mouse strains. "It was way overdue," says project leader Molly Bogue.

In addition, the project should help researchers planning knockout or mutagenesis projects to choose the most appropriate strain to work with. Attempts to investigate genes that may be involved in narrowing of the arteries, for instance, may best be done in mice with naturally high blood cholesterol.

### Character assassination

But even with such information, knockout experiments will continue to throw up mice that show no obvious phenotype. Many mouse genes belong to families whose functions overlap, and this 'redundancy' may mean that a clear phenotype only emerges when two or more genes are removed.

For example, knocking out the mouse gene *Uch-L3*, which codes an enzyme involved in breaking down regulatory, misfolded or damaged proteins, creates mice that are indistinguishable from their genetically intact relatives. But mice also lacking the related gene *Uch-L1* develop walking difficulties, paralysis and eventually die early from degeneration of nerve cells in the spinal cord<sup>8</sup>.

Although such examples do get reported, many knockout experiments in which no phenotype could be found never see the light of day. "A lot of those things you don't hear about," says Barbara Knowles, director of research at the Jackson Laboratory. To address the problem, the journal *Molecular and Cellular Biology* has, since 1999, given over a section to knockout and other mutant mice that seem perfectly normal.

Many of these animals might reveal their

phenotypes — if only researchers knew how to look for them. "I don't believe there is a single mouse that doesn't have a phenotype," says Mario Capecchi of the University of Utah in Salt Lake City, who shared a 2001 Lasker award for his pioneering work in developing the knockout technique. "We just aren't asking the right questions."

### Hidden traits

In some cases, a phenotype only becomes apparent when a mouse is exposed to particular environmental conditions. Mice made by Shoichi Kado, for example, developed cancer only when their intestines were colonized by bacteria. He and his colleagues at the Yakult Central Institute for Microbiological Research in Tokyo made mice lacking both the tumour suppressor gene *p53* and *TCR $\beta$* , a gene that regulates the immune system in the intestine. Mice reared in a standard, dirty mouse house developed intestinal tumours, they found, whereas those brought up in germ-free conditions had none<sup>9</sup>.

The large mouse-mutagenesis projects now under way are highlighting the need for more sensitive techniques for assessing phenotypes. Rather than disrupting an entire gene, as in knockouts, the mutagens used in these screens usually create changes in a single DNA base. This may favour the creation of correspondingly faint phenotypes, which can easily be missed. "Those subtle phenotypes will be the next wave of biology," predicts Mark Fishman of Harvard University, who works on both zebrafish and mouse mutants.

Some groups are improving the sensitivity of mutagenesis screens by tweaking their genetic strategy. Bill Stanford and his colleagues at the Samuel Lunenfeld Research Institute in Toronto, for instance, are crossing randomly mutagenized mice with animals that have a mutation in a single copy of the gene *c-kit*, which is involved in the production of blood cells and has been implicated in



P. FETTERS/HIMI

**I don't believe there is a single mouse that doesn't have a phenotype.**

Mario Capecchi

blood cancers. The team looks for mutations that suppress or enhance this phenotype by screening mice for altered levels of blood cells. "You will pick up things you would normally miss," says Stanford.

Sensibly, most of the big mouse-mutagenesis projects are concentrating on particular conditions, such as neurological or heart defects, and developing phenotypic screens that are optimized for the purpose. At Mount Sinai Hospital in Toronto, for instance, Janet Rossant hopes to use mouse-sized versions of medical imaging technology to identify mutants with defective growth or anatomy. The Mouse Imaging Centre will feature 16 magnetic resonance imaging scanners for high-resolution, high-throughput imaging of internal organs, and ultrasound to measure heartbeats from tiny embryos.

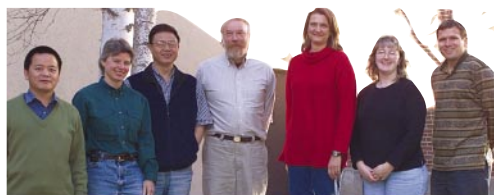
But even with such technological aids, phenotyping thousands of mutant mice is a formidable task. "We're going to have freezers filled with mutagenized sperm and not enough scientists to study the mice," Stanford predicts.

As experience with knockouts has shown, mouse genes often do not give up their secrets without a fight. In identifying the phenotype associated with an altered gene, and then working out why it only emerges against certain genetic backgrounds, there is no substitute for hard graft.

**Helen Pearson works in Nature's science writing team.**

1. Drab, M. *et al. Science* **293**, 2449–2452 (2001).
2. Nolen, P. M. *et al. Nature Genet.* **25**, 440–443 (2000).
3. Hrabé de Angelis, M. *et al. Nature Genet.* **25**, 444–447 (2000).
4. Threadgill, D. W. *et al. Science* **269**, 230–234 (1995).
5. Dietrich, W. F. *et al. Cell* **75**, 631–639 (1993).
6. MacPhee, M. *et al. Cell* **81**, 957–966 (1995).
7. Nadeau, J. H. *Nature Rev. Genet.* **2**, 165–174 (2001).
8. Kurihara, L. J., Kikuchi, T., Wada, K. & Tilghman, S. M. *Hum. Mol. Genet.* **10**, 1963–1970 (2001).
9. Kado, S. *et al. Cancer Res.* **61**, 2395–2398 (2001).

Mouse Phenome Project [www.jax.org/phenome](http://www.jax.org/phenome)



**Taking the strain: Molly Bogue (left, in red) is heading an effort at the Jackson Laboratory to track the characteristics of various lab mice.**



JACKSON LAB



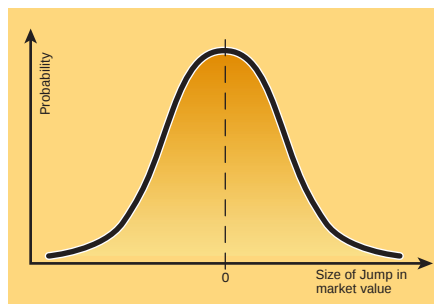
# The physics of the trading floor

Some physicists claim their modelling and data-analysis techniques can change the way we view stock markets. But mainstream economists have yet to be convinced, explains Mark Buchanan.

According to certain Internet sites, this week's lottery numbers can be predicted by identifying hidden patterns in previous draws. Probability theory, of course, suggests otherwise.

Standard economic theory similarly pours cold water on the idea that the behaviour of stock markets is affected by past market movements. Future behaviour of a market, say the textbooks, depends only on events in the real world, such as the profits and losses made by individual companies. So studying patterns in today's trading will not reveal the course of tomorrow's.

But many economists have long suspected that the textbooks are not telling the whole story. Some have argued that past trading does seem to have subtle effects on future fluctuations. Others in the new field of 'behavioural economics' suggest that the irrational psychology of investors lies behind these trends.



For many years, economists believed that market fluctuations could be described by the bell curve.



Trend setters: do investors unwittingly influence the future behaviour of stock markets?

Now these dissenters have gained some surprising allies: physicists intent on revealing hidden patterns in market behaviour. The jury is out on whether the field they have created — known as econophysics — will change mainstream economics, but their work is stirring debate. "Economics could use a little shaking up," says Blake LeBaron, an economist at Brandeis University in Waltham, Massachusetts. "There is no question that there is some interesting work coming out of econophysics."

## Telling tails

For much of the twentieth century, economists believed that the probability of a given change in the value of most markets — such as stock or foreign-currency exchanges — followed a pattern known as a bell curve (see left). This curve has two important properties: averaged across time, the most likely change is zero, and, because the curve tails off rapidly at extreme values, the probability of large fluctuations occurring is very low. Average stock prices may increase or decrease over the long-term, but day-to-day fluctuations seem to follow the bell curve.

Analysis of the data, however, shows that market behaviour is subtly different. In the 1960s, mathematician Benoit Mandelbrot, then at IBM's Thomas J. Watson Research Center in Yorktown Heights, New York, together with economist Eugene Fama of the University of Chicago, showed that markets

are better described by power-law distributions<sup>1,2</sup>. Power-law curves look superficially similar to bell curves, but their tails — the regions that cover large fluctuations — are different. Big jumps in market value are more common in power-law systems, giving rise to power-law curves' characteristic 'fat tails'.

But there are many different kinds of power-law curve, each of which has a differently shaped tail. The tail's shape is described using a parameter  $\nu$  — the higher the value of  $\nu$ , the faster the curve falls away, and the thinner the tail.

Over the past 20 years, several economists have shown that, in the case of fluctuations on the German stock exchange,  $\nu$  is greater than three<sup>3-5</sup>. This has important consequences, because statistical theory strongly suggests that systems with  $\nu$  greater than three cannot be random. For stock markets, this would mean that past performance can indeed offer clues about future fluctuations. Other work seemed to back this idea up. Economists already knew, for example, that large jumps in market value tend to cluster together over time<sup>6</sup>.

The results pointed to a new kind of market statistics. But with the precise value of  $\nu$  unknown, the exact nature of the statistics remained unclear. It is here that physicists found they could contribute. In 1995, Gene Stanley of Boston University, together with former colleague Rosario Mantegna, undertook what was then the most exhaustive study

AP

of market fluctuations ever. They analysed movements between 1984 and 1989 of the Standard & Poor's Index of the New York Stock Exchange (NYSE), which covers the 500 largest companies in the United States.

Coming from a physics background, Mantegna and Stanley were used to working with very large data sets. By examining nearly five million data points, they were able to back up the earlier work with a far larger body of data<sup>7</sup>. Together with colleagues from Boston University, Stanley went on to pin down the value of  $v$  at four for the Standard & Poor's index, as well as for the movements of the stocks of a 1,000 individual companies on the NYSE and other exchanges<sup>8</sup>. Other physicists have reported a similar value for foreign-exchange markets<sup>9</sup>. By using these truly gigantic data sets, physicists have been able to put much stronger constraints on the possible values of  $v$ , says Luís Amaral, one of Stanley's colleagues at Boston University.

### Deviant behaviour

But is any of this work, as the physicists involved believe, making a significant contribution to economic theory? Most economists remain sceptical. Deviations from random behaviour have been documented since the 1980s, they say. Economists already know, for example, that stocks that perform exceptionally well over one year tend to underperform relative to the market over the next<sup>10</sup>, suggesting that investors irrationally overvalue stocks that have risen in the recent past. "Most of the econophysics papers that I've seen are not particularly innovative," says Andrew Lo, an economist at the Massachusetts Institute of Technology. "In many cases they are reinventing things that economists have done many years ago."

But Doyne Farmer, a physicist at the Santa Fe Institute in New Mexico who has worked on problems in economics, defends the contributions of those in his field. Ten years ago, he says, trading firms characterized risk using bell-curve techniques. Many investment companies now use power-law statistics

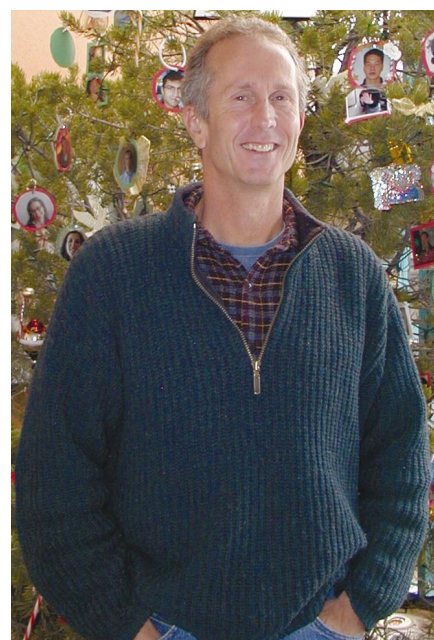
## The time has come for physicists to interact more closely with economists.

when analysing markets, and econophysicists feel this is in part because of their influence.

Physicists also point out that other techniques they have developed now form part of successful investment strategies. One ingredient in a good strategy is the selection of groups of assets that tend not to go up or down together. Diversifying holdings in this way helps investors to protect themselves against unexpected fluctuations in the market, as a drop in one group of assets tends to be cancelled out by an increase in the value of another. But this simple plan is surprisingly hard to implement, as underlying correlations between groups can be obscured by day-to-day price fluctuations.

Jean-Philippe Bouchaud is one of several physicists who have turned to random matrix theory — a mathematical technique developed to tackle problems in nuclear physics — to help solve the correlation problem. The theory can be used to identify the level of correlation that would be expected even if two stocks were not linked. This can then be used as a baseline for avoiding stocks that really do tend to vary in value together. In 1991, Bouchaud, who was then at the French Atomic Energy Commission in Paris, helped to establish Science & Finance, a Paris-based company that uses random matrix theory and other techniques to help to design investment portfolios.

Science and Finance is now used by several major investment companies. Despite this, many economists insist that physicists have done little more than analyse unusually large data sets. To silence their critics, econophysicists may need to make a more radical contribution to economics. A tantalizing



**Market value:** Doyne Farmer believes physicists are having an influence in the economic sphere.

link between behavioural economics and the statistics of power-law distributions may offer them the chance to do just that.

### Law enforcement

Power laws in physics are often seen in systems where the 'principle of universality' applies. The overall behaviour of such systems depends only weakly on the precise characters of the individual elements. Take the example of a substance that is heated until it is a mixture of a liquid and a gas. Under certain conditions, this system can be described using only the dimensions of the space it occupies, and the most basic geometrical details of how its constituent molecules interact with one another. All other aspects, such as the types of molecule involved and their precise size and mass, can be ignored. Consequently, crude models can capture important aspects of the system's behaviour.

As markets are also characterized by power laws, econophysicists hope to model them in a similar way. Markets are made up of independent traders, each of which assesses many economic factors before deciding whether to invest in a stock. But if the power-law conjecture is correct, market models need only consider a few crucial details of traders' behaviour, leaving out many economic variables. LeBaron has already used traders' behaviour as the basis for studies of markets<sup>11</sup>, but the power-law connection encouraged econophysicists to create simpler versions of his model. Initial results suggest that the new approach may have something to offer.

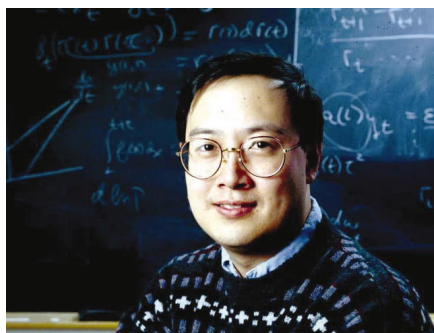
In 1997, for example, physicist Guido Caldarelli, then at the University of Manchester in England, together with researchers from the University of Fribourg in Switzerland,



**Called to account:** Luís Amaral (left) and Jean-Philippe Bouchaud believe that analytical techniques commonly used in physics can be deployed to understand and even predict market dynamics.







**Hard sell:** physicists must be more innovative to win economists' respect, says Andrew Lo.

created a computer model of a hypothetical market that contained just one stock. In this model, traders use simple mathematical strategies to predict future market changes, and base decisions to buy or sell the stock on these predictions. One trader might believe that recent trends will continue, for example, whereas another may bank on month-long cycles in market prices. Every so often, Caldarelli and his colleagues removed the least successful trader and introduced a new one with a random strategy in their place.

Despite the model's simplicity, and the use of arbitrary strategies, Caldarelli found that the stock's price behaved similarly to those on real markets — it followed an irregular pattern of rallies and crashes that were distributed according to a power law<sup>12</sup>. Neil Johnson, a physicist at the University of Oxford, claims to have used a variant of the same model to predict the movements of real markets, and says he hopes to commercialize his findings.

### Trade secrets

Some economists have been working along similar lines. In 1999, Thomas Lux, then at the University of Bonn in Germany, teamed up with Michele Marchesi, an electrical engineer at the University of Cagliari in Italy, to build a model in which the actions of one trader can directly influence the opinions and actions of others<sup>13</sup>. Lux and Marchesi divided traders into several different types. 'Optimists', for example, are more likely than 'pessimists' to buy if the market is rising.

The model shows the kind of behaviour predicted by fat-tailed distributions and hints at how big fluctuations might arise. Rising markets, for example, benefit the optimists, and so prompt some pessimists to switch strategies to become optimists. This increases overall investment, forces prices up and converts more pessimists into optimists. The result is a speculative bubble that eventually ends much as it started: a momentary drop in prices prompts a few optimists to become pessimists, and triggers a wave of conversions to the pessimistic strategy.

Power laws might also be active at even larger scales. In almost every nation, a small fraction of individuals has a large fraction of the wealth, and the distribution follows a

**A**dvocates admit the models are crude, but argue that they can provide valuable insights into financial systems.

simple power law. No one is sure exactly why this should be, but Bouchaud, together with Marc Mézard of the University of Paris South, has produced a model in which 1,000 individuals attempt to make money by trading among themselves or by investing<sup>14</sup>. The researchers found that the percentage return of investments — with the wealthy on average gaining or losing larger amounts than the less wealthy — inevitably leads to the power-law pattern.

Do these ideas hint at a new way of looking at economics? Physicists admit the models are crude, but argue that they provide valuable insights into financial systems. Others agree. Robert Axtell, a social scientist at the Brookings Institution in Washington, has successfully modelled the way in which companies are distributed by size in the United States<sup>15</sup>. "My model is very minimal," he says. "It is so spare as to seem quite unrealistic." Nevertheless, it produces intriguingly accurate results.

But so far, much of the work has failed to resonate with mainstream economists. Advocates of the models say that entrenched ideas may be to blame. "The idea of explaining statistical characteristics with behavioural models is alien to most economists," says Lux. "There is a hard core of conservative economists who will never accept these new ideas," Farmer adds.

The trading models go against the grain of standard economic thinking, which assumes that traders make judgements by considering company performance and the larger economic climate, and are not influenced by the actions of others. The models also ignore many details of the market that economists consider important. "Unless physicists are willing to make more of a commitment to understanding the intricacies of markets and economic dynamics, their work will not be taken very seriously by economists," says Lo.

Lo says that poor communication between the two camps is partly to blame. "There is very little common ground for exchanging ideas in a productive manner," he says. Farmer, meanwhile, admits that physicists need to catch up on what economists have already achieved. "The time is rapidly approaching when physicists who want to do serious work in finance need to interact more closely with economists," he says.

If econophysics is to make its mark, it seems there will have to be more mutual understanding. ■

**Mark Buchanan is a freelance writer in Livarot, France.**

1. Mandelbrot, B. B. *J. Bus.* **36**, 394–419 (1963).
2. Fama, E. F. *J. Bus.* **38**, 34–105 (1965).
3. Akgiray, V., Booth, G. G. & Loistl, O. *Allg. Stat. Arch.* **73**, 115–121 (1989).
4. Koedijk, K. G., Schafgans, M. M. A. & De Vries, C. G. *J. Int. Econ.* **29**, 93–108 (1990).
5. Lux, T. *Appl. Fin. Econ.* **6**, 463–475 (1996).
6. Engle, R. F. *Econometrica* **50**, 987–1008 (1982).
7. Mantegna, R. N. & Stanley, H. E. *Nature* **376**, 46–49 (1995).
8. Plerou, V., Gopikrishnan, P., Amaral, L. A. N., Meyer, M. & Stanley, H. E. *Phys. Rev. E* **60**, 6519–6529 (1999).
9. Müller, U. A., Dracorogna, M. M. & Pictet, O. V. in *A Practical Guide to Heavy Tails* (eds Adler, R. J., Feldman, R. E. & Taqqu, M. S.) 55–78 (Birkhäuser, Boston, 1998).
10. Schleifer, A. *Inefficient Markets: An Introduction to Behavioural Finance* (Oxford Univ. Press, Oxford, 2000).
11. Palmer, R. G. *et al. Physica D* **75**, 264–274 (1994).
12. Caldarelli, G., Marsili, M. & Zhang, Y.-C. *Europhys. Lett.* **40**, 479–484 (1997).
13. Lux, T. & Marchesi, M. *Nature* **397**, 498–500 (1998).
14. Bouchaud, J.-P. & Mézard, M. *Physica A* **282**, 536–545 (2000).
15. Axtell, R. L. *Science* **293**, 1818–1820 (2001).



**On the up:** physicists' models have proved surprisingly successful at predicting market trends.



# Intuition and inspiration made Gamow a star turn

The astronomer's rare insight meant that his guesses were often surprisingly accurate.

Sir — Horace Freeland Judson's forthright review of *Genes, Girls and Gamow* by James D. Watson (*Nature* **414**, 775–776; 2001) reminds me that in the summer of 1953, George Gamow did more than write to Watson suggesting how a linear sequence of four bases could result in different protein chains. He also delivered brilliant lectures at a now legendary astronomy summer school at the University of Michigan that I attended as his PhD student.

It is true that Gamow was funny and that he drank. It is also true that he was a brilliant scientist, devoted friend and concerned teacher, whose intuition exceeded that of any scientist I have known (see, for example, the Millennium Essay “The Big Bang and the genetic code” by Gino Segrè, *Nature* **404**, 437; 2000).

Walter Baade, probably the greatest

observational astronomer of the twentieth century, wrote in his book *Evolution of Stars and Galaxies* (Harvard University Press, 1963) the following: “I believe that George Gamow was the first to suggest the interpretation ... that is accepted today. ... Shortly after the publication of my paper on the two stellar populations, I received a typical message from Gamow on a postcard: ‘Please tell me where the lower branch of the color–magnitude diagram joins the main sequence, and I will tell you the age of your Population II stars.’” When Baade replied that not enough data existed, Gamow promptly answered: “With due respect to Schoenberg and Chandrasekhar, I have extrapolated the lower branch thus. O.K. Four to five billion years.” Baade concludes: “This was a guess, of course,

and that Gamow hit it so well was an accident, but his remarks really contained the whole story.” Gamow's remarks often contained the whole story, and his guesses were often correct.

This correspondence is the second I have submitted to *Nature*. The first, sent about 30 years ago, was to protest about a job advertisement for a radio astronomer, which stated that “equally qualified women will be paid  $x\%$  less than men”, where  $x$  was a number whose value (20?) I have forgotten. My contribution was not published because, the then-Editor wrote, in a year or so Australia was to put in place a law outlawing such discrimination.

**Vera C. Rubin**

Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road NW, Washington, DC 20015, USA

## Postdocs don't need reality to hit so hard

Sir — As a former French postdoc currently in ‘exile’ in England, mainly because of the depressing state of the current research system in France, I was outraged by Dr P.-L. Chau's Correspondence “Reality hits postdocs earlier in France” (*Nature* **414**, 582; 2001) responding to your News article (*Nature* **414**, 145; 2001). I can only assume that Chau is talking of another France, maybe on another planet, or that he mistakenly confuses the French tenured research system and the officially nonexistent (in terms of legal and social status) French postdoctoral system.

Postdocs in France are anything but civil servants, are typically at the mercy of their principal investigator, are often unable to conduct the research they want, and have to spend their time begging for short-term funding which, even if successful, barely supports a decent living. Of course one must not generalize, and progress is being made in improving the situation. But having experienced the system myself, and knowing numerous people who have suffered from it, or are still fighting to get a position in France, Chau's Correspondence is both erroneous and extremely depressing to all French postdocs.

Reality certainly hits the researcher earlier in France, but it couldn't hit any later than the PhD stage because by that time the person concerned is already

working abroad.

**Maryse Bailly**

Division of Cell Biology, Institute of Ophthalmology, University College London, 11–43 Bath Street, London EC1V 9EL, UK

## Talking about regeneration

Sir — In your timely, engaging and well-illustrated News Feature “The regeneration gap” (*Nature* **414**, 388–390; 2001), you report on the often-neglected invertebrate models of regeneration, such as planarians and hydra, and call for more attention to be paid to these invertebrates as models in the new field of regenerative medicine.

However, when referring to hydra and planarians, you describe stem cells as ‘dormant’ or ‘mobilized when needed’. Actually, the reverse is true. These cells, called neoblasts in planarians, are continuously mobilized in young and adult worms (see J. Bagnà, E. Saló & C. Auladell *Development* **107**, 77–86; 1989). Many planarians do not regenerate or do so poorly while bearing neoblasts with normal rates of cell renewal. Therefore, neoblasts are not ‘dormant’ cells for regeneration, but function to replace differentiated cells lost during daily wear and tear.

This process of renewal speeds up during regeneration, while the rates to different cell lineages change to adapt to the axial pattern remodelling of the

organism. Thus, terms such as ‘dormant’, ‘reserve’, or ‘mobilized when needed’ should be avoided when referring to stem cells in general and to planarian neoblasts in particular.

**Jaume Bagnà**

Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain

## Fruitful synthesis of science and fiction

Sir — In his Words essay “Where might it lead?” (*Nature* **414**, 399; 2001), Gregory Benford suggests how fiction can illuminate science. I agree that when a great writer weaves scientific concepts into his or her story, the reader is more likely to be drawn more deeply into the author's creation. I know of no other works of literature that marry the passions of science and unrequited love as eloquently as Umberto Eco's *The Island of the Day Before*.

Consider also some prose from Wallace Stegner's Pulitzer prizewinning 1971 novel *Angle of Repose*: “Many years later, when she reported that evening in her reminiscences, she was hearing the Doppler Effect of time, as I am now ... I think I hear the same tone, or tones, that she did: the sound of the future coming on for the girl of twenty-one, the darker sound of the past receding for the woman of eighty-four”, and later, “now I, Ahab, dismasted and with tunnel vision, seeing the back of my own

head through the curved lens of space-time". One can ponder the parallels between the significance of the title of the book and the concept of 'angle of repose' in the theory of Per Bak and collaborators describing the self-organized critical state of complex systems.

As an illustration of Benford's assertion that fiction informs science, Peter Atkins's *The Periodic Kingdom* is required reading in my honours chemistry class, as is a creative writing assignment based on this book's format (necessarily fiction). Although it is initially perceived as an odd assignment, I hope to infect web-surfing, computer-gaming students with an appreciation for literary science and fiction.

**Preston J. MacDougall**

Department of Chemistry, Box X101,  
Middle Tennessee State University, Murfreesboro,  
Tennessee 37132, USA

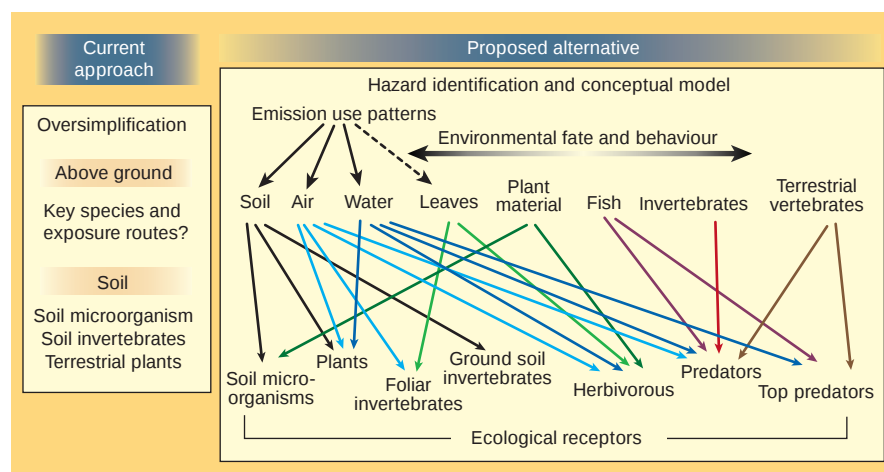
## Standardizing chemical risk assessment, at last

Sir — Industry is continually synthesizing new chemicals, the regulation of which requires evaluation of the potential danger for human health and the environment. Risk assessment is nowadays considered essential for making these decisions on a scientifically sound basis. Yet there are large data and conceptual gaps, which a new European Union (EU) white paper (policy document), *Strategies for a Future Chemicals Policy* ([http://europa.eu.int/eur-lex/en/com/wpr/2001/com2001\\_0088en01.pdf](http://europa.eu.int/eur-lex/en/com/wpr/2001/com2001_0088en01.pdf)) is attempting to redress.

The white paper is intended to clarify the definition and quantification of 'risk'; the margins of safety for description as 'low risk' (which currently show large differences for pesticides, veterinary drugs and industrial chemicals); and acceptability criteria (currently, the same concentration of the same chemical in soil or food items can be regarded as low or high risk depending on the EU guideline applied!).

Terrestrial ecosystems are of particular concern. In the past, ecotoxicologists have for various reasons focused on aquatic systems, so terrestrial risk assessments have been forced simply to apply the aquatic model to soils, or have focused on specific targets such as risk posed by agrochemical pesticides to birds, bees and beneficial arthropods.

An example of this confusion is demonstrated by the risk assessment of an insecticide that has an acute earthworm toxicity of 1 mg per kg (earthworm toxicity is a widely used measure, as earthworms are among the most sensitive soil-dwelling organisms). As things stand, such



**Comparison of the current and proposed approaches for the terrestrial ecological risk assessment of chemical substances.**

assessments can simultaneously conclude that:

(1) A farmer can use the insecticide as a plant-protection product without risk for soil organisms if the concentration in the soil does not reach 0.1 mg per kg.

(2) The same farmer cannot use the insecticide as a veterinary medicine on farm animals if this use could produce concentrations in the same soil higher than 0.01 mg per kg.

(3) Industry-related processes (for example, use of sludge for fertilizer) giving concentrations higher than 0.001 mg per kg in the same agricultural soil are classified as unacceptable risk, requiring risk refinement or risk reduction.

It is not clear whether confusing inconsistencies such as these originate from uncertainty, cost/benefit considerations, or the lack of scientific knowledge when the guidelines were set. Nevertheless, this inconsistency does not occur for aquatic risk assessment, where the rule is that a concentration 10 times below the chronic NOEC (highest concentration that does not produce effects) for the most sensitive aquatic species is acceptable. Any value above this trigger-point represents a potential risk.

Last year, the EU's Scientific Committee on Toxicology, Ecotoxicology and the Environment (CSTEE) reviewed the scientific basis of proper risk assessment on terrestrial ecosystems ([http://europa.eu.int/comm/food/fs/sc/sct/out83\\_en.pdf](http://europa.eu.int/comm/food/fs/sc/sct/out83_en.pdf)). The main weak points requiring further attention are the validation of model estimations with real monitoring data, and the development of a holistic approach for risk characterization. The new white paper requires much-needed simplified alternatives to speed up the assessment process. (At the current rate of progress, a comprehensive risk assessment of all chemicals currently on the market would take more than

1,000 years!). But, of course, simplifications should not jeopardize the best use of science. Now that the main research required to explore these alternatives has been identified by CSTEE, work can start, through the new EU chemicals strategy, while the Organisation for Economic Co-operation and Development and the United Nations Environment Programme can coordinate the extension of the approach beyond Europe.

We have compared the compartment-based approach developed in the 1980s and 1990s with the more holistic view now proposed by the CSTEE (see figure). Instead of the traditional two-compartment approach, each covering three taxonomic groups, the new proposal is to select key route-receptor interactions for each assessment. Targeted protocols can be referred to specific emission or use patterns, particular exposure routes or specific ecological receptors, all of which offer a large potential for covering regulatory needs.

The conceptual model for terrestrial ecosystems proposed in the figure can be used to assess the risks to humans and other species of exposure to chemicals on agricultural and other managed systems, for biological agents such as foot-and-mouth disease and bovine spongiform encephalopathy, and even radiation.

**J. V. Tarazona**

Spanish National Institute for Agriculture and Food Research and Technology (INIA), Carretera de La Coruña km 7, 28040 Madrid, Spain

Other signatories to this letter and members of the CSTEE Terrestrial Working Group:

**K. Hund** Fraunhofer Institute for Environmental Chemistry and Ecotoxicology, Germany

**T. Jäger** National Institute of Public Health and the Environment (RIVM), The Netherlands

**M. S-Salonen** University of Helsinki, Finland

**A. M. V. M. Soares** University of Aveiro, Portugal

**J. U. Skaare** National Veterinary Institute, Norway

**M. Vighi** University of Milano-Bicocca, Italy



# 1902 and all that

This year's offerings — calendars, kites and eruptions.

J. L. Heilbron and W. F. Bynum

Ours is a fortunate time indeed. For 1,000 years, folks could enjoy a palindromic year each decade. Then, with the turn of the first millennium, palindromic years came round only once a century. For the foreseeable future, only people alive around millennium turns will experience two palindromic years in one lifetime. *Fugit irreparabile tempus!*

In the suggestions that follow, we admit events occurring in preceding palindromic years as candidates for anniversarial celebration on an even footing with events that happened an integral number of 50 years ago. As usual, ¢ signifies 100 years, \$ a Nobel prize. We also admit, as usual, quarter-century anniversaries in the twentieth century to take account of the feverish pace of recent science. An example of the power and propriety of our procedure appears at the years 1552 (Eustachius his tube) and 1881 (Edward Talbot Ely his pioneering operation of otoplasty). On what other anniversarial principles would these aural landmarks qualify for celebration at the same time? Also, our principles link humans and hominids more closely than does the theory of evolution; thus 1991 (iceman found in the Alps), 1952 ± 1 (Pitdown man suspected as a fraud), 1952 ± 25 (Peking man discovered), 1902 (Neanderthal man reconstructed).

Despite our cornucopia of deserving events, the choice of anniversary of the year was easy: the decision of the British government to enter into calendrical communion with Europe 250 years ago. The British wisely rejected the Gregorian reform of 1582 on the likely ground that the world would not last long enough to justify the expense of conversion, and on the more reliable ground that all



The lost 11 days: Hogarth's *Election Entertainment* depicts the hostility at Britain's calendar reform.

changes are for the worse. When the world did not end and all the chancellor's tests were met, Britain dropped 11 days from the calendar in September 1752 and, for the first time in 170 years, agreed with Europe about the date. Some might find guarded comfort in this speedy precedent.



JOHN WEBB/BRIDGEMAN ART LIBRARY

## 1902 (1.0 ¢)

To begin at the top, we note that 100 years ago Arthur Edwin Kennelly and Oliver Heaviside independently deduced that the atmosphere must contain an electrified layer capable of reflecting the signals then recently sent into the ether by Guglielmo Marconi (\$1909). Léon-Philippe Teisserenc de Bort, proceeding upwards, thermometer in hand, observed that the atmosphere's temperature, after falling steadily for seven miles, remained constant to the greatest heights he could reach — which caused him to invent the words 'troposphere' for the lower region of change and 'stratosphere' for the upper region of layered consistency. The Kennelly–Heaviside layer lies above the stratosphere in the 'ionosphere'.

The great discovery of 1902 was the confirmation of the view, long held by Cassandras and cosmopolitans, that the world is falling apart. Ernest Rutherford and Frederick Soddy read from their data on the decline of activity of uranium X the news that the naturally occurring radioelements transform themselves by throwing off bits of



Moving on: Honold's 1902 magneto paved the way for the motor-car assembly line.

BETTMAAN/CORBIS





**Fighting anthrax in 1881: Pasteur performs his anthrax vaccination experiment at Pouilly-le-Fort.**

matter. The revelation of the suicidal tendency of atoms upset normal people as well as chemists. Where would it end? Rutherford (\$1908) and Soddy (\$1921) guessed helium: radioactive heavy elements struggle to become or to eject a gas. By 1902, physicists had become so inured to novelties that they digested exploding atoms with scarcely an eructation.

Not far behind the instability of matter in interest and novelty was the discovery that the organs of the body can communicate without a direct electrical connection between them. William Maddock Bayliss and Ernest Henry Starling announced that the pancreas could secrete its digestive liquid with all nerves severed between it and the intestines. They reasoned that the blood must carry a chemical trigger, which they called 'secretin' after its product; Starling later named such substances hormones from the Greek for 'arouse' or 'put in quick motion'.

That brings motor cars to mind. Gottlieb Honold got them off to a good start with a high-tension magneto; Louis Renault and Frederick Lanchester concentrated on an effective stop, one with a drum brake, the other with a disk. To keep track of things in between, the British firm of Thorpe and Salter introduced the speedometer, opti-

mistically graduated all the way to 35 miles per hour. James Edward Ransome's new motor-powered lawn mower did not need one. Nor did Hubert Cecil Booth's vacuum cleaner, the first of its kind — although it was drawn by swift horses, which made it inconvenient for indoor use. For a mess big enough, however, nothing could be more effective. England's Buckingham Palace and Windsor Castle each had one. The avant-garde could also have had an electric hair dryer, an electric typewriter and air-conditioning if they took in all the latest gadgets, and a teddy bear to console them if the power failed.

Among deaths in 1902 were those of Cato Maximilian Guldberg, Norwegian chemist and physicist, who worked out the law of mass action with his brother-in-law Peter Waage; Richard von Krafft-Ebing, Austrian psychiatrist, whose boring *Psychopathia sexualis* titillated generations of adolescents who could read its juicy bits in Latin; Johannes Wislicenus, German chemist, whose notion of geometrical isomerism inspired J. H. van't Hoff's (\$1901) theory of the tetrahedral carbon atom; and Rudolf Virchow, a paragon of a pathologist. Among births, that of Fritz Strassmann, the collaborator of Otto Hahn and Lise Meitner on the work that led to the discovery of nuclear fission, provides an opportunity for rescue from undeserved neglect.

### 1881 (1.21 ¢)

In this palindromic year, Joseph John Thomson (\$1906) calculated the electromagnetic mass associated with an electrified body in motion; Albert Abraham Michelson (born 1852, 1.5¢; \$1907) began his quest to detect the Earth's motion through the ether by use of his newly invented interferometer; and, with all-too-much contemporary relevance, Louis Pasteur dramatically demonstrated his anthrax vaccine at Pouilly-le-Fort.



**Going up in the world: Elisha Otis on his elevator.**

### 1852 (1.5 ¢)

Our pick for this drab successor to the year of the brilliant Crystal Palace is the safety elevator invented by the American Elisha Graves Otis, who demonstrated its effectiveness by cutting its cord while he rode in it. Only Henri Giffard's dirigible flight, the first ever, compares with Otis' short drop. Otherwise, the savant in search of a celebration must choose among Edward Franklin's proposal of something like valence; Hermann von Helmholtz's measurement of the velocity of nerve impulses; Rudolf Albert von Kölliker's elevation of sperm from a ferment to a cell; Henri-Victor Regnault's determination of absolute zero; and Edward Sabine's linkage of sunspot activity with disturbances in the Earth's magnetic field.

This is to omit the births of the American Society of Civil Engineers and of the brilliant Spanish neuroanatomist Santiago Ramón y Cajal (\$1906); and the deaths of Johan Gadolin, the Finnish chemist, discoverer of



**Clean sweep: but not yet house-trained.**



**Thomson: shown here in ordinary mass.**

yttrium and the eponym of gadolinium (1886, the first element named after an individual), and Louis Braille, inventor of the tactile alphabet for the blind.

### 1802 (2.0 ¢)

Two centuries ago, Anders Gustav Ekeberg found a new metal in a new mineral from the prolific region around Ytterby in Sweden. He named it tantalum because its oxide refused to combine with acid even when fully immersed, just as the unfortunate Tantalus could not drink the water in which he stood. Jean-Baptiste Lamarck and Gottfried Treviranus independently coined another enduring neologism that year, 'biology'; and Gerardus Johannes Mulder, inventor of the word 'protein', was born.

In 1802 Humphry Davy, then beginning his brilliant career in chemistry at the Royal Institution, demonstrated that metal slips could be heated to incandescence by an electric current; Thomas Wedgwood, the son of the potter, and Davy presented the institution with the negative results of their efforts to render permanent images cast by sunlight on a paper coated with silver nitrate; and Thomas Young exposed the institution's audiences to his principle of superposition, whereby he endeavoured to explain the diffraction of light and the colours of thin plates. It would be premature, however, to celebrate Young's celebrated double-slit experiment, now regarded as a definitive confirmation of the wave motion of light and, suitably modified, of matter. We must wait until 2007.

It remains to announce the deaths of Franz Maria Ulrich Theodor Aepinus, who worked out a theory of electricity based on forces decreasing with the distance, independently of Henry Cavendish; Erasmus Darwin, physician, botanist, poet, inventor, friend of the Wedgwoods and grandfather of Charles; and Marie-François-Xavier Bichat (born in the palindromic year 1771), who made general anatomy the study of tissues.

### 1771 (2.31 ¢)

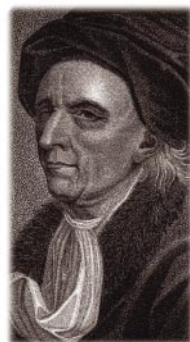
In 1771 Charles Messier published the first instalment of his catalogue of nebulae, undertaken to help distinguish them from comets, which he delighted to hunt. The last of his catalogues, for 1784, brought his nebular count to 103. Then William Herschel took up the game with his huge reflector, and by 1800 had found some 2,000. That gigantism in instrumentation can pay is not a discovery of the twentieth century.

### 1752 (2.5 ¢)

While Britain and its colonies, preparing to convert to the Gregorian calendar, worried about how to reckon rents and contracts, Benjamin Franklin was flying a kite. The sport was doubly successful: he showed the similarity of lightning and electrical dis-

charges and survived to advertise the experiment. He had reason to underrate its danger. A few months previously, an old French dragoon, directed by the Comte du Buffon, had performed Franklin's test. The dragoon brought his knuckle close to an insulated iron rod projecting above the roof of a small guardhouse during a storm. A priest attended — to estimate the duration of the sparks by reciting the rosary and to render last rites if Franklin had miscalculated. Neither Franklin nor the dragoon received a lightning stroke.

Jumping over the disagreeable experiments on the digestive juices of hawks that René de Réaumur performed this year, we arrive at the pure theory of Leonhard Euler, who, 250 years ago, deduced the formula



1752: mathematician Leonhard Euler.

$V - E + F = 2$  relating the number of vertices  $V$ , edges  $E$ , and faces  $F$  of a regular polyhedron. It follows from this innocent formula that five, and only five, such polyhedra — the number recognized by Plato — can exist. For this insight Euler received nothing special. Not so his fellow mathematician Johann Tobias Mayer, whose lunar tables, published

in 1752 along with directions for finding the longitude at sea, won him a share of the great prize offered by the British Board of Longitude for a suitable solution to this problem of the age. Acting with the swiftness of English calendar reform, the board paid Mayer after his death. Incidentally, the method of lunars was taught in Nathaniel Bowditch's *New American Practical Navigator* of 1802 (2.0 ¢).

### 1702 (3.0 ¢)

David Gregory, a Scottish mathematician and pedagogue, achieved the notable feat of enlisting the support of both Newton's enemy John Flamsteed and Newton himself in his successful bid for the Savilian professorship of astronomy at Oxford. He comes to mind for his *Astronomiae physicae et geometricae elementa* (1702), which transmits Newton's belief that the ancients knew the inverse-square law of gravity. Thomas Savery's 'miner's friend' also looked backwards and forwards. It described Savery's application of seventeenth-century ideas of steam and atmospheric pressure to raising water and anticipated the more practical and powerful steam engine invented by Thomas Newcomen. The first recorded Newcomen engine dates from 1712 (2.9 ¢).

### 1661 (3.41 ¢)

There is only one choice, Robert Boyle's *Skeptical chymist*, which argued against the systems of Aristotle and Paracelsus, showed

chemists how to use the mechanical philosophy to design experiments and express results, and exhibited some new and useful classes of substances.

### 1652 (3.5 ¢)

It was a Scandinavian lymph year. Thomas Bartholin of Copenhagen published his *De lacteis thoracis*, demonstrating the lymphatic system in humans, and Olof Rudbeck of Uppsala showed the same to Queen Christina, using a dog. He (not the dog) died in 1702 (3.0 ¢). Our preference for a 3.5-¢ celebration is the birth of Wilhelm Homberg, chemist, alchemist and cosmopolitan, who published his *magnum opus*, beginning in 1702 (3.0 ¢), under the modest title *Essais de chimie*.

### 1602 (4.0 ¢)

We can offer Tycho Brahe's posthumous *Astronomiae instauratae progymnasmata*, which gives the positions of 777 stars and a description of the supernova of 1572 (4.3 ¢).

### 1551/2 (4.5/4.51 ¢)

Early products of the English Reformation include the author of the first book in English on melancholy (the English disease) (Timothy Bright, b.1551), and the publication of the first monographic description of the English malady, the sweat (1552, by John Caius).

On the European continent, Erasmus Reinhold, professor of higher mathematics at the protestant University of Wittenberg, and his colleague in the chair of lower mathematics, Georg Joachim Rheticus, were the leading early exponents of the heliocentric system of the Catholic Copernicus. Rheticus arranged for Copernicus' manuscript to be printed; Reinhold translated the theory into tables for astronomical observations. These *Tabulae prutenicae* (1551), named after their patron, Duke Albrecht of Prussia, replaced the Alfonsine Tables named after Alfonso X of Castile and Léon (acceded 1252, 7.5 ¢).

### 1502 (5.0 ¢)

This year marks the 500th anniversary of the return of that engaging liar Amerigo Vespucci



Stjerneborg: Tycho Brahe's observatory on the island of Hven.





**Mystery voyage: Vespucci seeking the New World.**

(died 1512, 4.9 €) to Lisbon from Brazil. His racy descriptions of the sexual licence and cannibal cuisine of the Guaraní caused a sensation that no doubt encouraged him to concoct an account of his alleged voyage to the New World in 1497. This account, written in 1504, claimed priority over Columbus in the discovery of what is now known as South America. A German cartographer who accepted the fictional claim baptised the new world America. Columbus rated Vespucci the only honest ship chandler he knew.

#### 1991 (0.11 €)

It was a murky year. Smoke from the oil wells in Kuwait ignited by Iraqi troops could be detected north of Turkey, west of the Sudan and as far as China. Mount Pinatubo erupted on the island of Luzon, the largest eruption of the twentieth century, massively polluting the atmosphere and damaging the ozone layer.

More worthy of celebration were the first imaging of buckyballs; the detection of volcanoes (perhaps active) on Venus; the orbiting of the 17-tonne Gamma-Ray Observatory; and the marketing of a new Cray supercomputer capable of 16 gigaflops.

#### 1952 (0.5 €)

As a small dividend on the first explosion of a thermonuclear device, Glen Seaborg (\$1951) and his group at Berkeley identified a new element among the products in the

blast, which they named einsteinium in continuation of the practice begun with gadolinium (see 1852). Other nuclear events full of promise were the completion of the first breeder reactor in the United States, and the provocation (by human error) of the first reactor accident, in Chalk River, Canada. In purer physics, the Brookhaven Cosmotron came online at over 1 GeV, and Donald Glaser (\$1960) observed the tracks of cosmic rays in his new bubble chamber.

Stanley Lloyd Miller and Harold Urey (\$1934) exposed pure water, hydrogen, ammonia and methane to an electric spark, creating amino acids — which, had they been allowed another 10 billion years of lab time, might have formed sentient creatures.

New drugs and active therapeutic substances appeared with wonderful regularity, recognized sometimes by the Nobel institution, which awarded its medicine or physiology prize to Selman Waksman for streptomycin. The tuberculosis armamentarium was further enhanced by isoniazid, by Edward Heinrich Robitzek and his colleagues, leading many to assume that tuberculosis would soon be history. Other potent drugs of the year included erythromycin, reserpine, 6-mercaptopurine and chlorpromazine. Cortisone was synthesized and aldosterone isolated. When Charles Sherrington (\$1932) died, aged 95, it seemed a safe bet that his would soon be an ordinary lifespan.

We note, too, the death of Chaim Weizmann, Russian-born president of Israel, the Zionist leader who obtained from Lord Balfour the famous declaration in favour of a Jewish homeland. Weizmann had access to high places through *Clostridium acetobutylicum* — a compliant bacterium whose properties he had investigated before the war — which breaks down starches into ethanol, acetone and butanol. During the First World War, Weizmann industrialized the process to make acetone in great quantities for the explosive cordite and created an obligation later translated into support of a Zionist state.

#### 1952 ± 25 (0.5 ± 0.25 €)

*Plus ça change.* In 1927, Charles Lindberg made the first transatlantic solo flight, from New York to Paris, in 33.5 hours; in 1977, Bryan Allen flew the *Gossamer Condor*, a man-powered aircraft, 1.35 miles. In 1927, the first transatlantic telephone call came by electrical cable from London to New York; 1977 saw the first practical use of optical cables. In 1927, Georges Lemaître announced his hypothesis of the cosmic egg, a forerunner of the Big Bang; 50 years later, Alan Guth looked behind the egg to the inflationary Universe. Seventy-five years ago, H. J. Muller (\$1946) grilled genes with X-rays; half a century later Phillip Sharp and Richard Roberts (\$1993) independently split them and revealed a lot of useless information.

Charles Elton's *Animal Ecology* (1927)

came at the beginning of a long career devoted to studying (among other things) animals adapted to hostile environments. The most notable may be the giant molluscs, tube worms and bacteria that share a sulphur-based metabolism in deep-sea vents off the Galapagos Islands, discovered by John Corliss and Robert Ballard in 1977. To finish, 1927 saw the foundation of the Verein für Raumschiffahrt (space travel club), which numbered Wernher von Braun (died 1977) among its members; and 1977, the launching of Voyagers 1 and 2 towards the outer planets.

*Yin and yang.* If we could be sure about the time and place, we would attend this year's big party for the principle of uncertainty, discovered or invented by Werner Heisenberg (\$1932) 75 years ago. Heisenberg's formulation of the principle displeased Niels Bohr (\$1922) for exaggerating particles over waves; and later in 1927 he tabled his own interpretation of quantum mechanics, which buys even-handedness at some cost in intelligibility. Bohr could point to experiments of 1927 which demonstrated the wave-like properties of electrons (Clinton Davisson and, independently, George Thomson; \$1937). His 'complementarity', which places the task of physics in correct and communicable description of phenomena, has at least this in common with Percy Bridgman's (\$1946) contemporary *The Logic of Modern Physics* (1927), that it allows into physical theory only entities for which a procedure of measurement can be stipulated.

The paradoxes and complementarities observable in 1977 include the last known case of smallpox, in Somalia, and the diagnosis of two cases of Kaposi's sarcoma, in New York City; the launch, by Stephen Wozniak and Steve Jobs, of their Apple II, the first personal computer for everyone, and the creation of Microsoft by Paul Allen and Bill Gates. ■

J. L. Heilbron is at Worcester College, Oxford OX1 2HB, UK. W. F. Bynum is at the Wellcome Trust Centre for the History of Medicine, University College London, 24 Eversholt Street, London NW1 1AD, UK.



**Apple's roots: Jobs (left) and Wozniak (right).**



**1991: Mount Pinatubo erupts on Luzon.**

# The measure of a Victorian polymath

Pulling together the strands of Francis Galton's legacy to modern biology.

**Sir Francis Galton: From African Exploration to the Birth of Eugenics**

by Nicholas Wright Gillham

Oxford University Press: 2001. 432 pp.

£22.50, \$35

Garland E. Allen

Despite Francis Galton's importance to scientific life in the post-darwinian period, only two biographies of this polymath have been written — a voluminous *Life, Letters and Labours* by his devoted protégé Karl Pearson between 1914 and 1930, and a more manageable attempt by D. W. Forrest in 1974 (*Francis Galton: The Life and Work of a Victorian Genius*). Arguing that Pearson's work is too large and unwieldy (with which I agree) and that Forrest's presents a picture of Galton that is too disembodied (with which I do not agree), Nicholas Gillham attempts to improve on both by focusing on Galton as “a creature of flesh and blood”. Although this new biography clearly relates Galton's work more closely to modern concerns than does Forrest's, it is less clear that it is significantly better in capturing Galton's personality and psychology. Like those before him, Gillham has encountered the problems caused by Galton's secrecy about his innermost feelings, and by the daunting task of integrating the many facets of his extraordinary career into a unified whole. The result is a highly readable and interesting — but still somewhat disjointed — portrait of Galton as scientist and man.

The book begins with Galton's geographical explorations. Restless after leaving Cambridge with only a pass degree, Galton went off to the Nile in 1845, initiating a love affair with Africa that produced an expedition to Ovamboland in Namibia, financed and led by Galton, during 1850–52. Galton's penchant for measuring things found its first systematic outlet here in his recordings of latitudes, longitudes and altitudes, collections of meteorological data, and map drawing. Particularly interesting are Gillham's descriptions of Galton's day-to-day activities, his encounters with the local inhabitants and his cunning ways of wooing them into supplying guides and provisions. Although elitist and racist, Galton distinguished between those Africans he found more intelligent and civilized and those he considered true “savages”.

Technical articles published in the *Journal of the Royal Geographical Society* and three popular travel books in the 1850s won him instant recognition as one of the leading geographers of his age. Married in 1853, Galton settled into the life of an indepen-



**Tools of his trade:** Galton's anthropometric laboratory was equipped and maintained at his own expense.

dently wealthy London intellectual.

Most of the second half of the book is devoted to Galton's various intellectual pursuits: these included the study of hereditary talent and character, the use of twin studies to separate the effects of nature and nurture, pioneering work in statistics (including the methods of calculating correlation coefficients and regression analysis), analysis of fingerprints as a means of identification, the invention of the composite photograph as an anthropometric tool, and finally the promulgation of eugenics. One of the strengths of Gillham's book is the clarity with which he explains these disparate areas of study. Although Galton may appear as a dilettante in some ways, it is also clear that he took his work and mission in life seriously, and was ahead of his time in introducing statistics and measurement into biological investigations.

The intermingling of Galton's personal and scientific sides works best in the early part of the book. Here the vivid personal details — the journey up the Nile, Galton's methods for organizing the Namibian expedition, and his ongoing public clashes back in London with Henry Morton Stanley (the journalist who went in search of Dr Livingstone) to name just a few — make lively reading. The personal details become more distracting in the latter half of the book, however, largely because Galton's life after marriage and “settling down” was banal. The personal details provide little insight into either Galton

as a person or his intellectual work.

A similar lack of integration runs through the book as a whole, even when it is dealing explicitly with Galton's intellectual pursuits. Although Galton was famed for his constant metrical analyses of everything (from “boredom” to “the efficacy of prayer”), we do not learn where this obsession originated and how it became a guiding principle for so much of his scientific work. Mathematical thinking did not come easily to him, and he had to work hard at Cambridge, helped by a tutor, to prepare for exams. More probing here might have provided some interesting insights into the origin of biometrics, an important area of modern biology.

The book's final two chapters deal with Galton's initiation of the field of eugenics, and his collaboration with mathematician Karl Pearson. As a biologist himself, Gillham is particularly interested in the origin of Galton's eugenics, and what eugenics has come to mean to us today. Here, the integration of the personal and the intellectual comes across clearly — Galton's patrician sensibilities were disturbed by the “degeneration” of English society that he saw around him. Eugenics, a term coined by Galton himself, was intended to mean “well or truly born”. Galton favoured both positive eugenics (encouraging highly able people to have many children) and negative eugenics (discouraging or preventing lower-grade people from having any children), although he was



not in favour of coercive legislation. Galton's vision of eugenics was more abstract and benign than that which emerged in the United States and Germany in the first four decades of the twentieth century. As Galton's final obsession, eugenics became the crowning point of his long career, and the work for which he is probably most frequently remembered.

Admirable though it is in many ways, the book has flaws. Gillham tries at times to cast Galton's ideas in present-day terms that make them sound more modern than they really are. For example, Galton's statement that the "quasi-independent units which the body is made up of have a separate origin, or germ" is not really equivalent, as Gillham claims, to "genes specifying individual proteins"; nor is Galton's explanation of reversion equivalent to recessive mendelian genetics. I understand the author's desire to make apparently opaque and confused theories comprehensible to the modern reader, but such policy does a disservice to both the reader and to Galton. That Galton found little of interest in Mendel's theory after its rediscovery in 1900, for example, attests to the very different conceptual understanding of heredity under which he and his contemporaries, including Darwin, were working.

All in all, this is a highly readable and valuable contribution to the fields of both biography and the history of science. Gillham has done a creditable job in placing Galton in his Victorian social and intellectual environment, and in showing where he was constricted by — and able to transcend — his own time and place. ■

Garland E. Allen is in the Department of Biology, Washington University in St Louis, Campus Box 1137, St Louis, Missouri 63130, USA.

## State-of-the-art oceanography

**Ocean Circulation & Climate: Observing and Modelling the Global Ocean**

edited by Gerold Siedler, John Church & John Gould

*Academic: 2001. 715 pp. \$99.95, £65.95*

Detlef Quadfasel

Ocean science has made a big leap forward during the past decade. Developments outside the field itself, such as increasing computer power and new measurement

technology, have driven much of the advance. The global ocean can now be observed from space and *in situ* with autonomous moored and floating devices, and marine climate variability can be modelled on this basis. The societal interaction in oceanography has also improved, due primarily to the World Ocean Circulation Experiment (WOCE), carried out during the 1990s, which was the first truly global scientific undertaking to attempt to understand (and eventually predict) the role played by the oceans in the Earth's climate. The effort was global not only in terms of scientific coverage and participation by scientists, but also in its data-sharing policy. The development of modern electronic communications also greatly contributed to its success. *Ocean Circulation & Climate*, an account of the results of this experiment, is not just a cruise report but provides a comprehensive view into the state of the art of blue-ocean research.

This modern view of the ocean is communicated in some 30 articles, covering different technical, regional, scientific and political-philosophical aspects of WOCE and the ocean. The outline of the book largely follows the organizational structure of the experiment with its working groups and panels, but there are also discussion chapters integrating the results in a broader sense. This strategy has the advantage that individual subjects are covered by experts in each field. On the other hand, its disadvantage is that it has the character of a conference proceedings report, with a concomitant variability in scope and quality of the individual contributions.

WOCE has undoubtedly proved (again) that the ocean is turbulent and highly variable, and that the historical view of a large-scale and quasi-steady dynamic balance is inadequate, if not totally misleading. Variability exists on all timescales, and has been quantified in the surface drifter and float programmes, with the moored current meter records and the satellite altimetry data, for the first time providing global statistics. Small-scale processes such as mixing along a density gradient are explored and shown to be associated with strong spatial and temporal variability. Its influence on the large-scale circulation is just beginning to emerge, but may lead to a radical change in ocean modelling.

WOCE has also shown that the large-scale circulation itself can undergo rapid changes in response to atmospheric forcing — much faster than previously thought. The intense observational programme in the North Atlantic has documented this coupling through the processes of convection and overflows, and has provided a basis for the study of oceanic variability associated with the Arctic or North Atlantic Oscillations. Here the interaction between the



## Monstrous manifestations

Beware this black dragonfish (*Idiacanthus antrostomus*), which uses a phosphorescent lure to attract its prey. The image is taken from *Creatures of the Deep: In Search of the Sea's*

"Monsters" and the World They Live In by Erich Hoyt (Firefly, \$40). The book explores the domain far below the waves and reveals that monsters "are alive and well and living in the deep".



## Down on the shore

American avocets (*Recurvirostra americana*) can be identified by their upwardly curving beaks and contrasting plumage. The numbers of these gregarious shoreline birds are under threat from loss of their wetland

habitat in the western United States. The avocets are just one of the 72 species of bird described in *Common Coastal Birds of Florida and the Caribbean* by David W. Nellis (Pineapple Press, \$21.95).

atmosphere and the deep ocean is fast, and ranges from years to decades rather than centuries or longer. High-quality hydrographic observations of water masses over the globe have identified the variability in the ocean's interior and allowed quantification of the impact of varying atmospheric forcing. A series of chapters is devoted to the analysis of hydrographic data, which constituted the core of WOCE. From these, one can conclude that the initial WOCE goal of compiling a coherent global oceanic description has led to better documentation of the ocean's variability. The development of models for predicting associated climate change is well under way.

A third message conveyed in the book is probably as old as physics itself: that adequate observations are a prerequisite to test, accept or refute hypotheses, theories and models. Climate dynamics are highly nonlinear, and the predictive skill of any model can only be judged if adequate and long-term observations are available. The search for methods for determining long-term changes in the ocean circulation has been, and remains, the second goal of the programme. Many technical advances have been made, and the ideas about new and cheap measurement strategies are described in several of the book's chapters. Much of this has already been implemented in follow-up projects under the umbrella of the World Climate Research Programme. The necessity of extending the time series of observations relating to climate is reiterated throughout, and it can only be hoped that politicians and

science administrators in funding agencies will take the opportunity to read this book and learn about the ocean.

Who else is a potential reader? The review character makes *Ocean Circulation & Climate* attractive for graduate students in Earth sciences and climate researchers in general, and of course the WOCE community itself. As an overview containing almost 100 pages of up-to-date references, this book will be an invaluable tool for those engaged in climate research and teaching. It is certainly not a book that one reads from beginning to end, but one that will be picked up readily when questions about ocean circulation arise.

Detlef Quadfasel is in the Department of Geophysics, Niels Bohr Institute for Astronomy, Physics and Geophysics, University of Copenhagen, Juliane Maries Vej 30, 2100 Copenhagen, Denmark.

## More on the oceans

### The Oceans and Rapid Climate Change: Past, Present and Future

edited by Dan Seidov, Bernd J. Haupt & Mark Maslin  
American Geophysical Union, \$60

### Wind Stress Over the Ocean

edited by Ian S. F. Jones & Yoshiaki Toba  
Cambridge University Press, £50, \$80

### Great Waters: An Atlantic Passage

by Deborah Cramer  
W. W. Norton, £22, \$27.95

## Name that plant

### Guide to Standard Floras of the World

by David G. Frodin  
Cambridge University Press: 2001. 1,100 pp.  
£150, \$240

P. F. Stevens

Here, in a mere 1,000 pages or so, we have a guide to the plants of the world, past and present. David Frodin focuses on standard floras — a catalogue of plants of a defined area with descriptions, but excluding lists of weeds — past and present, and his is by far the most comprehensive treatment of the subject available.

There are, clearly, large gaps in our knowledge about the world's plants. For example, if you want to know about the plants to be found in Baluchistan, there is only I. H. Burkill's *A Working List of the Flowering Plants of Baluchistan*, published almost a century ago. Moreover, as Frodin notes, both his book and floras as a discipline in general are at a crossroads, and it is not clear what form either will take in the future. Frodin's book was current to 2000, but it is easy to imagine an online version being continually updated.

The *Guide to Standard Floras* will be invaluable as a library resource. Three introductory chapters lead to the body of the work, a geographical listing of floras. The lists of plants in the major geographical areas identified by Frodin — for example, divisions, areas often continental in size — are preceded by informative summaries of the state of our knowledge of the plants in them. The book is made navigable by several indexes — geographical names, the numbers referring to Frodin's geographical system (not pages), and authors of floras, although it is somewhat confusing because Frodin makes references to both page numbers and geography. In general, the indexes are complete, although there is no mention of the island of Bougainville, and not one of the 14 authors of the remarkable *Flora da Reserva Ducke* (one of a new breed of floras) is mentioned; this has more than 20,000 photographs for some 2,200 species, and emphasizes the use of vegetative features in identification.

The introductory chapters clearly indicate the complex institutional, national and professional constraints that have shaped the way floras have developed, and also demonstrate how web technology is straining the conventional format of the flora. In the past, floras have included identification aids, detailed descriptions of plants, lists of localities, specimens cited, literature,



guides to flowering times, maps, illustrations, and the like, all variously bundled up. Writing a flora, or working on a flora team, has in the past been a major part of the training of systematists, although in most European and North American institutions this is no longer the case. Flora writing has also been the life's work of many a senior systematist — but floras have lost some of whatever cachet they once had. Frodin does not discuss the current hegemony or predominance of molecular work. Concentrating solely on such work further decreases the likelihood of young systematists picking up a working knowledge of plants of an area. Of course, at the same time, our understanding of species-level variation can be revolutionized by molecular-based phylogenographical studies.

As Frodin mentions, there has been a long-standing tension between the writers of floras and the local botanist. Thus in the nineteenth century, many flora writers did not simply hold to a broad, herbarium-based species concept, but they deliberately dismissed the findings of local botanists, regardless of whether they were German academics or colonists. It is hardly surprising that even in the United States, 'biosystematic' information has only belatedly been incorporated into written floras. However, both the local botanist and someone studying local patterns of variation bring an additional perspective to understanding variation that can be absent in the work of a monographer. Add a molecular component, and our understanding of diversity, particularly in the tropics, may need revision.

Indeed, one can easily see how the web and well-organized databases may help us to cut loose from the shackles of conventional floras; most of us do not want to see listings of specimens, literature, lengthy descriptions, and so forth, all together. Similarly, those who want to place their flora in the context of the most recent phylogenetic studies will simply opt to refer to the Angiosperm Phylogeny Group's consensus classification, whereas others will prefer an alphabetical arrangement. These and other preferences could be catered for in a web-based system. Such developments could even bring some *de facto* uniformity and standards to botanical subdisciplines that are noted for their insane idiosyncracies (often hidden under the refrain of 'the expert always knows best'). More importantly, the different needs of users of botanical information could be addressed, and the study of plants stimulated at the same time. ■

P. F. Stevens is in the Department of Biology, University of Missouri, St Louis, Missouri 63121, and at the Missouri Botanical Garden, Monsanto Research Center, St Louis, Missouri 63166, USA.

## Science in culture

### Cool for nerds

**SEED — a new magazine.**

Josette Chen

Adam Bly wants people to stop thinking that scientists are frizzy-haired men in lab coats, with pocket protectors and bubbling purple flasks. As founder and editor-in-chief of a new, ultra-cool 'science couture' magazine called *SEED*, he sees his mission as redefining the way science is seen in popular culture.

His inspiration was fired when, at the tender age of 18, he attended the 1999 UNESCO World Conference on Science in Budapest. He perceived there a discrepancy between the popular science media — as portrayed by the likes of *Scientific American* and *Discover* — and the evolution of science in popular culture. The media needed to evolve, he concluded.

So, with this in mind, Bly built up a media group, called SEED Group, with its headquarters in Montreal and with the magazine as its main focus, funded by private investors in Canada from the worlds of business, science and fashion.

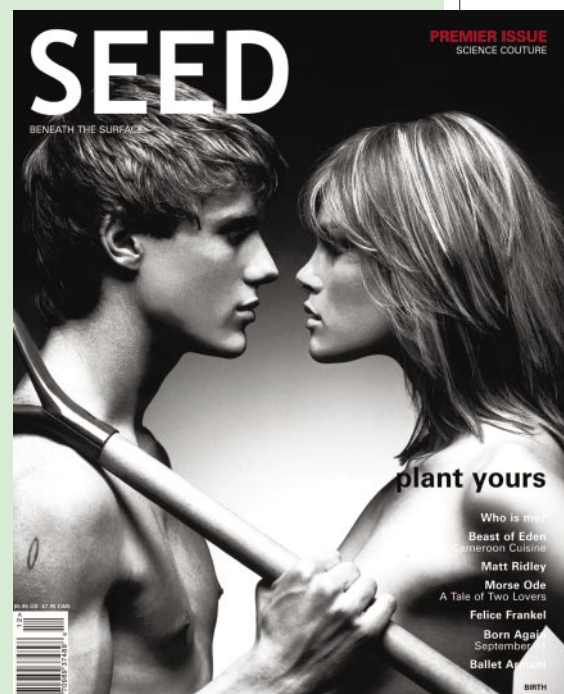
*SEED* looks like a cross between *Vogue* and that monthly bible of cool, *Wallpaper*. Its first issue was launched in November, with a distribution of 100,000 across North America. Six issues a year are planned, along with a doubling of distribution.

The 176-page premiere issue, with a theme of birth, is filled with evocative imagery interspersed with factoids and a few longer essays, all in some way pertaining to science or technology. Topics range from the seductive powers of chocolate, to the bushmeat trade, to the demise of Starlab in Brussels.

The contributor list is impressive, leading with popular-science writer Matt Ridley — he of *Genome* fame — on the wonder and strangeness of human birth. The work of Felice Frankel, an MIT researcher and science photographer, and Karl Ammann, environmental photographer, also grace the pages. In between, male and female models in states of undress decorate fluffy sound-bites on consciousness and the language of lovers. Even the advertisements are sexy; there are no be-goggled biochemists showing off their favourite PCR in these pages — instead, pouty-lipped models promote fashion designers and alcohol distillers.

The written science is low level, although *SEED*'s description of why the sky is blue — "blue waves agitate the elements of the sky, imparting their zeal" — can at least be said to depart from the typical scientist's mumblings about Rayleigh scattering and one over lambda to the fourth.

Priced below both *Nature* and *Vogue*, *SEED* is aimed at the science professional and "image-conscious science aficionados". "We are targeting a sophisticated and affluent demographic," says Bly. "The reader is a scientist who is looking for entertainment, to diversify his or her interests, and to learn how to influence culture." Even the



**The magazine seeks to redefine the way science is seen in popular culture.**

more conservative of *Nature*'s readers are image-conscious, he believes. Any scientist who wants to get his experiment on the cover of *Nature* (or *Science*) shows his desire to produce aesthetically pleasing scientific imagery.

In the future there will be new sections devoted to the current icons of science — people who are part of Bly's vision of redefining science — the role the media plays in conveying science to the public, and the role of science in business and politics and vice versa. Top writers are continuing to sign up. According to Bly, the next issue will have contributions ranging from Nobel laureates and Pulitzer prizewinners. Why? Ridley, for one, strongly believes in the cause of making science more hip, and more appealing to young people: "Science has no need to be wordy with huge woggles of text."

Bly set up a media group to leverage his vision into other ventures. He says that the next two years will be devoted to the magazine, but that they are also looking into book publishing and television projects in 2003.

At least with the magazine, scientists everywhere at last have a venue where they may find advice on de-frizzing their hair and replacing their scruffy lab coats with the latest Hugo Boss fashions. ■

Josette Chen is a postdoctoral fellow at the Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada.

European distribution of *SEED* will begin in March.

# Geology of mankind

Paul J. Crutzen

For the past three centuries, the effects of humans on the global environment have escalated. Because of these anthropogenic emissions of carbon dioxide, global climate may depart significantly from natural behaviour for many millennia to come. It seems appropriate to assign the term 'Anthropocene' to the present, in many ways human-dominated, geological epoch, supplementing the Holocene — the warm period of the past 10–12 millennia. The Anthropocene could be said to have started in the latter part of the eighteenth century, when analyses of air trapped in polar ice showed the beginning of growing global concentrations of carbon dioxide and methane. This date also happens to coincide with James Watt's design of the steam engine in 1784.

Mankind's growing influence on the environment was recognized as long ago as 1873, when the Italian geologist Antonio Stoppani spoke about a "new telluric force which in power and universality may be compared to the greater forces of earth,"

referring to the "anthropozoic era". And in 1926, V. I. Vernadsky acknowledged the increasing impact of mankind: "The direction in which the processes of evolution must proceed, namely towards increasing consciousness and thought, and forms having greater and greater influence on their surroundings." Teilhard de Chardin and Vernadsky used the term 'noösphere' — the 'world of thought' — to mark the growing role of human brain-power in shaping its own future and environment.

The rapid expansion of mankind in numbers and per capita exploitation of Earth's resources has continued apace. During the past three centuries, the human population has increased tenfold to more than 6 billion and is expected to reach 10 billion in this century. The methane-producing cattle population has risen to 1.4 billion. About 30–50% of the planet's land surface is exploited by humans. Tropical rainforests disappear at a fast pace, releasing carbon dioxide and strongly increasing species extinction. Dam building and river diversion have become commonplace. More than half of all accessible fresh water is used by mankind. Fisheries remove more than 25% of the primary production in upwelling ocean regions and 35% in the temperate continental shelf. Energy use has grown 16-fold during the twentieth century, causing 160 million tonnes of atmospheric sulphur dioxide emissions per year, more than twice the sum of its natural emissions. More nitrogen fertilizer is applied in agriculture than is fixed naturally in all terrestrial ecosystems; nitric oxide production by the burning of fossil fuel and biomass also overrides natural emissions. Fossil-fuel burning and agriculture have caused substantial increases in the concentrations of 'greenhouse' gases — carbon dioxide by 30% and methane by more than 100% — reaching their highest levels over the past 400 millennia, with more to follow.

So far, these effects have largely been caused by only 25% of the world population. The consequences are, among others, acid precipitation, photochemical 'smog' and climate warming. Hence, according to the latest estimates by the Intergovernmental Panel on Climate Change (IPCC), the Earth will warm by 1.4–5.8 °C during this century.

Many toxic substances are released into the environment, even some that are not toxic at all but nevertheless have severely damaging effects, for example the chlorofluorocarbons that caused the Antarctic 'ozone hole' (and which are now regulated). Things could have become much worse: the

## The Anthropocene

*The Anthropocene could be said to have started in the late eighteenth century, when analyses of air trapped in polar ice showed the beginning of growing global concentrations of carbon dioxide and methane.*

ozone-destroying properties of the halogens have been studied since the mid-1970s. If it had turned out that chlorine behaved chemically like bromine, the ozone hole would by then have been a global, year-round phenomenon, not just an event of the Antarctic spring. More by luck than by wisdom, this catastrophic situation did not develop.

Unless there is a global catastrophe — a meteorite impact, a world war or a pandemic — mankind will remain a major environmental force for many millennia. A daunting task lies ahead for scientists and engineers to guide society towards environmentally sustainable management during the era of the Anthropocene. This will require appropriate human behaviour at all scales, and may well involve internationally accepted, large-scale geo-engineering projects, for instance to 'optimize' climate. At this stage, however, we are still largely treading on *terra incognita*.

Paul J. Crutzen is at the Max Planck Institute for Chemistry, PO Box 3060, D-55020 Mainz, Germany, and the Scripps Institution of Oceanography, University of California, San Diego, 9500 Gillman Drive, La Jolla, California 92093-7452, USA.

### FURTHER READING

- Marsh, G. P. *Man and Nature* (1864). (Reprinted as *The Earth as Modified by Human Action* (Belknap Press, Cambridge, Massachusetts, 1965)).
- Crutzen, P. J. & Stoermer, E. F. *IGBP Newsletter* **41** (Royal Swedish Academy of Sciences, Stockholm, 2000).
- Clark, W. C. & Munn, R. E. (eds) *Sustainable Development of the Biosphere* Ch. 1 (Cambridge Univ. Press, Cambridge, 1986).
- Vernadsky, V. I. *The Biosphere* (translated and annotated version from the original of 1926) (Springer, New York, 1998).
- Turner, B. L. *et al.* *The Earth as Transformed by Human Action* (Cambridge Univ. Press, Cambridge, 1990).
- McNeill, J. R. *Something New Under the Sun: An Environmental History of the Twentieth-Century World* (W. W. Norton, New York, 2000).
- Houghton, J. T. *et al.* (eds) *Climate Change 2001: The Scientific Basis* (Cambridge Univ. Press, Cambridge, 2001).
- Berger, A. & Loutre, M.-F. *C. R. Acad. Sci. Paris* **323** (IIA), 1–16 (1996).
- Schellnhuber, H. J. *Nature* **402**, C19–C23 (1999).





# Breaking up a superfluid

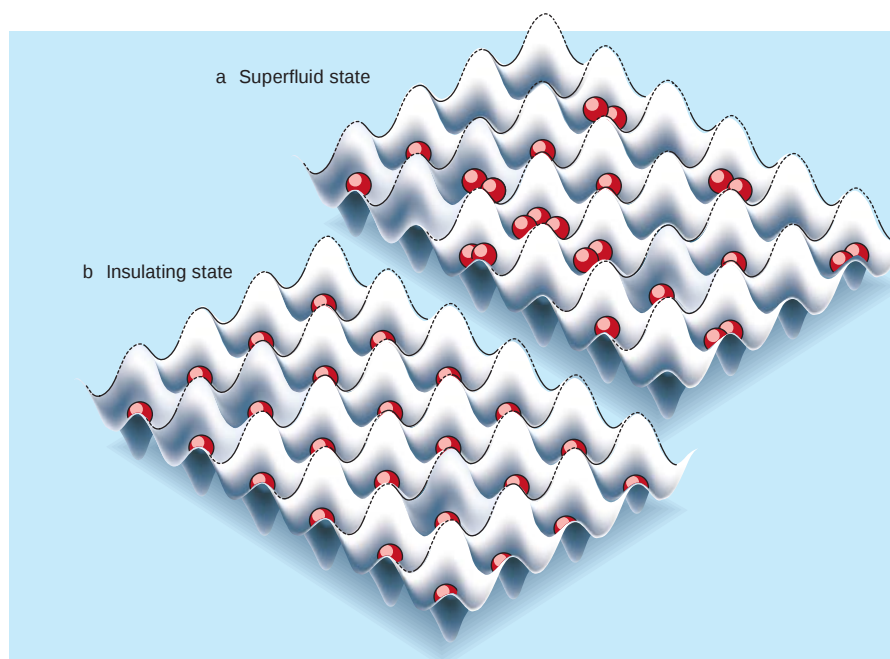
Henk T. C. Stoof

Ultracold atoms held in a three-dimensional pattern by a web of light beams can now be switched from a superfluid to an insulating state. This achievement may be useful for performing quantum computations.

At extremely low temperatures of less than one-hundred-millionth of a degree above absolute zero (10 nanokelvin), the atoms in a rubidium gas essentially all join into a single quantum state to form a Bose–Einstein condensate. In such a condensate the atoms can flow without friction, and so the gas is a superfluid. When this superfluid is placed in an energy landscape consisting of high-energy mountains and low-energy valleys there is initially little effect (Fig. 1a). In particular, the superfluid nature of the gas does not change and the atoms still move freely from one valley to the next. But when the mountains become just a little too high the atoms suddenly lose their freedom, and each atom is trapped in a single valley (Fig. 1b). This surprising behaviour — switching a quantum gas from a superfluid to an insulating phase — has been demonstrated for the first time by Greiner *et al.*<sup>1</sup>, as they report on page 39 of this issue.

The experiment by Greiner *et al.*<sup>1</sup> represents a landmark in the history of Bose–Einstein condensed gases, for several reasons. The creation of a Bose–Einstein condensed gas of rubidium atoms in a magnetic trap is challenging in itself, and the first researchers<sup>2</sup> to do so in 1995 were recently rewarded with the 2001 Nobel prize for physics. Several laboratories around the world can now create Bose–Einstein condensates almost routinely, but Greiner *et al.*<sup>1</sup> also needed a periodic energy landscape for the atoms. This is most easily achieved using the electric field of a laser beam or, more precisely, several criss-crossing laser beams, which create a light-wave interference pattern known as an optical lattice.

Optical lattices have also been around since the mid-1990s, and a standard three-dimensional optical lattice is made using four laser beams, so the resulting lattice is less sensitive to instabilities in the lasers. But Greiner *et al.*<sup>1</sup> used six laser beams to create an optical lattice with an extra twist. With the additional laser beams they can make an energy landscape in which the distance between the valleys (and mountains) in all three directions is exactly the same — it is a perfect cubic lattice. Such experimental wizardry allows the authors to demonstrate convincingly the transition of the ultracold atoms from superfluid to insulating behaviour and back again. To confirm the state of the quantum gas at any instant they simply



**Figure 1** A quantum phase transition in an ultracold gas. By using a web of laser beams to create an energy landscape of mountains and valleys (an optical lattice), Greiner *et al.*<sup>1</sup> can reversibly switch a gas of rubidium atoms from a superfluid to an insulating phase. **a**, At a temperature of 10 nanokelvin or less the rubidium atoms share the same quantum state and are in a superfluid phase, in which they can move freely between valleys. **b**, By increasing the intensity of the laser beams in the optical lattice, the researchers force the gas into an insulating phase, in which each atom is trapped in an individual valley. Such control is vital to most proposals for building a quantum computer.

switch off the magnetic trap and the optical lattice, and then take an image of the gas cloud after it has expanded for 15 milliseconds. Only when the gas is a superfluid do they observe a beautiful interference pattern.

For a physicist, the observation of this transition is exciting because it is solely the result of the Heisenberg uncertainty principle. Such ‘quantum phase transitions’ have attracted a lot of attention in recent years, because they are fundamentally different from their more familiar ‘classical’ counterparts that are driven by thermal rather than quantum fluctuations. For example, true quantum phase transitions occur only at a temperature of absolute zero. So, in principle, thermal fluctuations will still have a pronounced effect on the transition<sup>3</sup>, even at temperatures as low as 10 nanokelvin. In the case of a quantum gas in an optical lattice, understanding the effect of thermal fluctuations remains an important challenge. In this respect, theorists are particularly interested in the speed of sound in the gas near the transition point.

In its most familiar form, Heisenberg’s uncertainty principle says that knowing precisely the position of a particle prevents you knowing precisely its momentum, and vice versa. What you gain on one hand you lose on the other. The uncertainty principle at work in Greiner *et al.*’s experiment<sup>1</sup> forbids simultaneously knowing the number of atoms in a certain valley, and the phase of the condensate’s ‘wave function’ in that same valley. Because all the atoms in a Bose–Einstein condensate occupy a single quantum state, they are described by a quantum wave function whose phase is exactly the same in each valley. As a result, when the gas cloud is in the superfluid state the number of atoms in each valley can vary considerably. In the insulating phase the situation is reversed: the number of atoms in each valley is now fixed, so the phase of the wave function changes randomly from one valley to the next.

This particular quantum phase transition was first studied theoretically by Fisher *et al.*<sup>4</sup> in the context of granular superconductors and Josephson-junction arrays. It

was also predicted to occur for a quantum gas in an optical lattice by Jaksch *et al.*<sup>5</sup>. Their proposal has several advantages, as demonstrated by Greiner *et al.*, because there are no lattice imperfections (disorder) so the physics of the quantum phase transition occurs in its most ideal form. Another advantage of using a quantum gas in an optical lattice is that the height of the mountains in the energy landscape can easily be changed by varying the intensity of the laser fields. This makes it possible to switch back and forth between superfluid and insulating behaviour. For granular superconductors and Josephson-junction arrays it is essentially impossible to achieve the same control over the energy landscape because creating a new landscape requires a new sample. So one cannot actually see the superfluid–insulator transition taking place in a single system.

A tantalizing application for the ideal array of single atoms created in the insulating phase is quantum computing. Every rubidium atom has a magnetic moment and so has

two internal states that may serve as the 0 and 1 of a quantum bit. Because there are a large number of rubidium atoms in the optical lattice, they can act as a memory for a quantum computer. Moreover, if there are two such memories that can be moved relative to each other, we can even make use of the interactions between the atoms to perform a quantum computation<sup>6</sup>. The first step towards this exciting goal has now been taken. ■

Henk T. C. Stoof is in the Institute of Theoretical Physics, Utrecht University, Leuvenlaan 4, 3584 CE Utrecht, The Netherlands.  
e-mail: h.t.c.stoof@phys.uu.nl

1. Greiner, M., Mandel, O., Esslinger, T., Hänsch, T. W. & Bloch, I. *Nature* **415**, 39–44 (2002).
2. Anderson, M. H., Ensher, J. R., Matthews, M. R., Wieman, C. E. & Cornell, E. A. *Science* **269**, 198–201 (1995).
3. Sachdev, S. *Quantum Phase Transitions* (Cambridge Univ. Press, 2001).
4. Fisher, M. P. A., Weichman, P. B., Grinstein, G. & Fisher, D. S. *Phys. Rev. B* **40**, 546–570 (1989).
5. Jaksch, D., Bruder, C., Cirac, J. I., Gardiner, C. W. & Zoller, P. *Phys. Rev. Lett.* **81**, 3108–3111 (1998).
6. Jaksch, D., Briegel, H.-J., Cirac, J. I., Gardiner, C. W. & Zoller, P. *Phys. Rev. Lett.* **82**, 1975–1978 (1999).

## Ageing

# The price of tumour suppression?

Gerardo Ferbeyre and Scott W. Lowe

The p53 protein works to suppress cancer, so one might think that bumping up the levels of this protein would be a good idea. But this isn't so — mice with too much p53 age prematurely.

The p53 gene is often touted as 'the most frequently mutated gene in human cancer'. As a result, a great deal is known about it. When fully functional, p53 works to suppress the development of tumours, and current thinking suggests that it does so by affecting how cells respond to damage<sup>1</sup>. p53 can be activated by many stresses, such as breaks in DNA, and in turn induces responses that keep cell numbers down, including cell death (apoptosis) and permanent cell-division arrest (senescence; Fig. 1a). When p53 is mutated, cells cannot respond correctly to stress and are predisposed to becoming cancerous.

Indeed, mice that lack p53 develop normally but rapidly succumb to cancers<sup>2</sup>, implying that, at the whole-organism level, p53 acts purely as a tumour suppressor (Fig. 1b). But on page 45 of this issue, Tyler *et al.*<sup>3</sup> identify an unexpected aspect of p53 biology. They show that mice engineered to have high p53 activity are resistant to tumours — but age prematurely. Their results raise the shocking possibility that ageing may be a side effect of the natural safeguards that protect us from cancer.

Tyler *et al.*'s findings<sup>3</sup> were serendipitous. They were trying to engineer mice with a particular mutation in p53, and instead produced animals in which a large portion

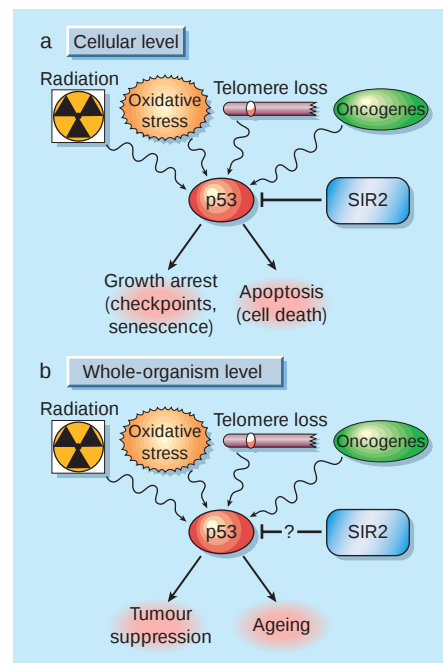
at one end of the p53 gene (the 5' end) was missing. The remaining portion, which they call the *m* version of the gene, is just the 3' end and encodes a shortened protein. Mice with one normal copy of the p53 gene and one copy of the *m* version were resistant to tumour formation, but even relatively young animals resembled old normal mice. By contrast, mice with one non-functional p53 gene and the *m* gene were indistinguishable from mice that lack p53 altogether. So it seems that a normal p53 gene is needed for the *m* gene to have an effect, hinting that the protein encoded by the *m* gene steps up the activity of normal p53. This seems remarkable, yet is consistent with the observation<sup>4</sup> that short peptides encoded by the 3' end of the p53 gene can activate normal p53.

It is not yet certain that the features of ageing are only due to the mutant p53 gene. The authors could not detect the truncated protein in the mutant mice, and at least one essential gene near the p53 gene was also deleted. Still, a second p53-mutant strain also shows signs of premature ageing<sup>3</sup>. Of course, the definitive experiment — showing that loss of p53 increases longevity — is impossible, because p53 and cancer are so intertwined.

Tyler *et al.*'s results might simply reflect a highly abnormal, diseased state. But they

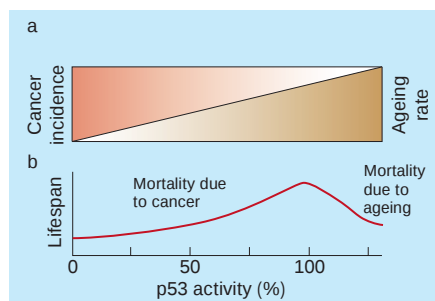
may also reveal a role for p53 in normal ageing. What might this role be? The evolutionary theory of ageing suggests that organisms invest resources to ensure reproductive success early in life, at the expense of longevity. It is not clear how this works at the molecular level. One model proposes that loss of the normal mechanisms for quelling gene expression leads to genes being expressed incorrectly and, ultimately, to ageing. So, enforced expression of the gene-silencer SIR2 extends lifespan in yeast and worms<sup>5</sup>. Another model holds that non-reproductive cells have a clock that monitors how often they have divided (based on the lengths of telomeres, protein–DNA complexes that cap chromosome ends and erode gradually with each cell division). Short telomeres eventually trigger senescence. Indeed, mice with critically short telomeres age prematurely<sup>6</sup>.

In a third model, cellular damage leads to mutations or general cellular decline, ultimately producing the spectrum of age-related characteristics<sup>7</sup>. Support for this idea comes from the fact that radiation increases



**Figure 1 The p53 system and ageing.** a, At the cellular level several stresses can activate p53, inducing checkpoints in the cell-division cycle, permanent cell-division arrest (senescence) and cell death. Suppression of p53 activity prevents these responses, predisposing cells to becoming cancerous. b, At the whole-organism level, p53 activation results in a lower cancer incidence. But Tyler *et al.*<sup>3</sup> show that p53 can also promote ageing. The SIR2 protein, which increases longevity in experimental organisms such as yeast, can associate with p53 and suppress its activity in cultured cells (a). It is not known whether this process contributes to whole-organism ageing (b). Tyler *et al.*'s results suggest that aspects of ageing may arise as by-products of p53's tumour-suppressor activity.





**Figure 2** Balancing cancer and ageing. **a**, Increases in p53 activity reduce the incidence of cancer but increase the ageing rate. Conversely, decreases in p53 activity increase cancer incidence but may decrease the rate of ageing. **b**, As a consequence, the influence of p53 on lifespan may result from a delicate balance between its antitumour and pro-ageing effects, such that too little p53 increases mortality from cancer whereas too much p53 increases mortality from ageing.

mutations and promotes premature ageing in mice<sup>8</sup>. Tyner *et al.*'s results add an interesting twist to this hypothesis — they suggest that ageing may be due in part to cellular responses to damage, and not damage per se (Fig. 1b). These responses protect most people from cancer during their reproductive years, but they may come at a price.

All of these models can be loosely linked to p53. The SIR2 protein can suppress p53 activity in mammalian cells<sup>9,10</sup>. Cellular responses to telomere malfunction involve p53 (ref. 11), and a lack of p53 can counteract some of the effects of a lack of telomerase (an enzyme that regenerates telomeres) in mice<sup>12</sup>. Finally, doses of radiation that promote premature ageing in mice activate p53 efficiently<sup>13</sup>. Despite these connections, p53 has been on the periphery of ageing research — until now.

The new work<sup>3</sup> also has other implications. The ageing of whole organisms has been linked to cellular senescence *in vitro*, a process that also involves p53 and was initially linked to the finite lifespan of cultured human fibroblast cells<sup>14</sup>. Senescent cells remain metabolically active but cannot proliferate; they also show changes in gene expression that could produce alterations at the tissue level<sup>15</sup>. 'Replicative' senescence is triggered by telomere erosion and can be prevented by telomerase. But the same events can also be produced in response to, for example, DNA damage, oxidative stress and suboptimal cell-culture conditions<sup>14</sup>; telomerase has no effect here. Confusion over the exact process has led to a semantic debate, but it is clear that senescence parallels apoptosis as a cellular response to stress. These facts, coupled with the new results<sup>3</sup>, suggest that tissue ageing *in vivo* results from many factors, not just telomere attrition as has been suggested.

It might seem paradoxical that overactive p53 suppresses cancer but promotes ageing, given that the incidence of cancer usually

increases with age. But the problem can be resolved by the fact that cancer results from the malfunctioning of p53 in single cells, whereas ageing involves a tissue-wide process. So, cells with inactive p53 ultimately shorten lifespan because cancer develops. Conversely, cells with abnormally high p53 activity do not contribute to cancer, but instead undergo cell death or senescence. With time, these changes may compromise tissue physiology, shortening lifespan through ageing. So p53 activity must be tightly controlled to balance a predisposition to cancer (too little p53) and premature ageing (too much p53; Fig. 2).

Finally, Tyner *et al.*'s work<sup>3</sup> could have ramifications for understanding and treating human diseases. For example, p53 might contribute to premature ageing syndromes or age-related disorders in humans. And although longevity is complex and involves many p53-independent factors, it is conceivable that variation in lifespan is influenced by variation in p53's response to cellular damage. The results also raise the disturbing possibility that the DNA-damaging drugs used to treat cancer in young people might prompt p53 into action and accelerate age-related disorders later on. This is a testable hypothesis.

## Cosmology

# A baryometer is back

Corinne Charbonnel

The usefulness of helium-3 as a probe of the early Universe has been in doubt. A rethink of stellar theory and new observational data put those doubts to rest.

**D**o we live in an open Universe that will expand forever, or in a closed Universe whose expansion will eventually reverse? Part of the answer can be found by accurately estimating the amount of ordinary — baryonic — matter in the Universe. One way of reaching such an estimate stems from one of the predictions of Big Bang theory: that production of light nuclei (deuterium, helium-3, helium-4 and lithium-7) occurred within the first seconds of the Universe, in a process known as Big Bang nucleosynthesis. A precise measurement of cosmological baryon density requires determination of the primordial abundances of light nuclei, and for those estimates to be consistent with each other<sup>1</sup>.

On page 54 of this issue<sup>2</sup>, Bania, Rood and Balser report the first reliable assessment of the primordial abundance of <sup>3</sup>He. Their result is based on two decades of radio observations of star-forming H II regions and planetary nebulae in our Galaxy, the Milky Way, and on theoretical developments in the field of stellar evolution. Beyond the observational challenges, which in themselves are consider-

able, the prospect alone underscores the need for less toxic anticancer drugs. But perhaps the most provocative implication is that, if p53 is involved in ageing, then drug-related approaches to interfere with this process may come at a price — the disruption of our natural mechanisms for keeping cancer at bay. ■

Gerardo Ferbeyre is in the Department of Biochemistry, Université de Montréal, Montréal, Quebec, Canada, H3C 3J7.

e-mail: g.ferbeyre@umontreal.ca

Scott W. Lowe is at the Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA.

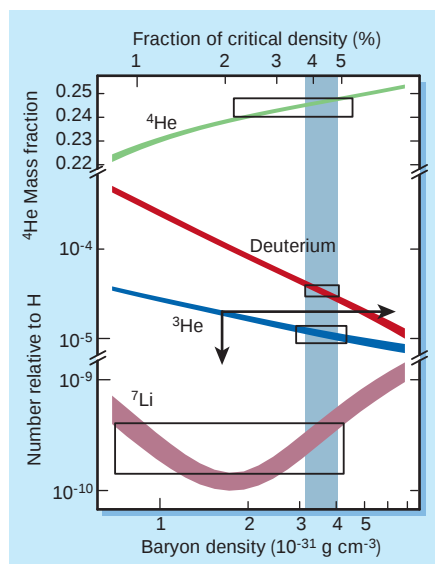
e-mail: lowe@cshl.edu

1. Vogelstein, B., Lane, D. & Levine, A. J. *Nature* **408**, 307–310 (2000).
2. Donehower, L. A. *et al.* *Nature* **356**, 215–221 (1992).
3. Tyner, S. D. *et al.* *Nature* **415**, 45–53 (2002).
4. Hupp, T. R., Sparks, A. & Lane, D. P. *Cell* **83**, 237–245 (1995).
5. Guarente, L. & Kenyon, C. *Nature* **408**, 255–262 (2000).
6. Rudolph, K. L. *et al.* *Cell* **96**, 701–712 (1999).
7. Szilard, L. *Proc. Natl Acad. Sci. USA* **45**, 30–45 (1959).
8. Curtis, H. J. *Science* **141**, 686–694 (1963).
9. Luo, J. *et al.* *Cell* **107**, 137–148 (2001).
10. Vaziri, H. *et al.* *Cell* **107**, 149–159 (2001).
11. Karlseder, J., Broccoli, D., Dai, Y., Hardy, S. & de Lange, T. *Science* **283**, 1321–1325 (1999).
12. Chin, L. *et al.* *Cell* **97**, 527–538 (1999).
13. Komarov, P. G. *et al.* *Science* **285**, 1733–1737 (1999).
14. Campisi, J. *Trends Cell Biol.* **11**, S27–S31 (2001).
15. Shelton, D. N., Chang, E., Whittier, P. S., Choi, D. & Funk, W. D. *Curr. Biol.* **9**, 939–945 (1999).

able, the main difficulty in using Big Bang nucleosynthesis as a cosmological probe lies in departures from the primordial abundances. Almost everywhere, the chemical composition of the Universe has been modified by processes such as stellar nucleosynthesis and cosmic-ray collisions. To infer the primordial abundances of the light elements from those measured, this chemical evolution has to be understood and quantified.

According to the classical theory of stellar evolution, formulated in the early 1970s, low-mass stars such as our Sun should be producing large amounts of <sup>3</sup>He. One dying star, the planetary nebula NGC3242, does indeed do this. NGC3242, which is slightly more massive than the Sun, previously synthesized fresh elements in its interior, and is ejecting them into the interstellar medium. Among those elements is <sup>3</sup>He, and it is being produced in the amounts predicted. In consequence, the expectation has been that the amount of <sup>3</sup>He in the Galaxy would increase over time<sup>3</sup>.

Helium-3 can be observed only in relatively young objects in the Milky Way, such as the Sun, the local interstellar cloud<sup>4</sup>, a



**Figure 1 Big Bang nucleosynthesis predictions for the primordial nuclei. The boxes indicate the observational estimates for the primordial abundances of  $^4\text{He}$  (observed in galactic H II regions), deuterium (inferred from the absorption spectra of high-redshift quasars) and  $^7\text{Li}$  (deduced from the surface of the oldest stars in our Galaxy). For  $^3\text{He}$ , the horizontal arrow points to the upper limit of the abundance, relative to hydrogen. The vertical arrow points down to the lower limit of the baryon density inferred from the  $^3\text{He}$  upper limit, assuming a Hubble constant of  $65 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . The horizontal lines of the black box over the dark-blue  $^3\text{He}$  theoretical band indicates the range of abundances derived from the H II region S209; the vertical lines show the allowed range of baryon densities, which agree with the range of densities determined from deuterium abundances as shown by the vertical blue band. Modified from ref. 9.**

few planetary nebulae, and H II regions. The H II regions, which have just formed out of matter that has undergone 12 billion years of chemical evolution, should be particularly rich in  $^3\text{He}$ . But when Bania and collaborators started to look at them in 1984, 'the  $^3\text{He}$  problem' emerged: there was no evidence of  $^3\text{He}$  enrichment having occurred. In consequence, the usefulness of  $^3\text{He}$  as an indicator of baryon density — as a 'cosmological baryometer' — was thrown into doubt.

But ways have been found out of this quandary, both through theory and now observations. The lack of increase in  $^3\text{He}$  abundance can be accounted for if most low-mass stars consume most of their  $^3\text{He}$  before it can be emitted into the interstellar medium. This requires physical processes which the classical theory of stellar evolution did not take into account, but which have been revealed by anomalous carbon isotopic ratios observed at the surface of old stars. These anomalies could be explained by stellar rotation: by inducing internal mixing, rotation simultaneously leads to the destruction of  $^3\text{He}$  inside stars, so it is not emitted into the interstellar medium. About 10% of stars would be unaffected, however, NGC3242 being one of them. This alternative view<sup>5</sup> results in a much reduced  $^3\text{He}$  contribution from low-mass stars compared with that in the classical scheme of events.

Bania and collaborators<sup>2</sup> now report progress on the observational side of things. They have tenaciously continued to sample H II regions, increasing the number covered and so probing a large part of the Galactic disk. Obtaining these data is quite a feat, as it has involved accumulating measurements of extremely weak spectral lines of  $^3\text{He}$  at centimetre wavelength with the National Radio Astronomy Observatory's 140-foot telescope. Deriving an estimate of  $^3\text{He}$  abundance from the resulting spectra is also no easy job, as it requires technically demanding modelling of the complex structure of the sources. One of

the authors' refinements was to select for their sample H II regions with relatively simple structure, for which it was easier to achieve accurate determinations of  $^3\text{He}$  abundance.

Bania and colleagues find that, wherever H II regions are in the Galaxy, they all have a similarly low  $^3\text{He}$  content. Taken together with the theory that rotating stars will emit negligible  $^3\text{He}$  into the interstellar medium, this implies that the value of  $^3\text{He}$  in H II regions provides an upper limit to its primordial abundance, and a lower limit to the baryon density.

The upshot of Bania *et al.*'s 20 years of hard labour, then, is that they have rehabilitated  $^3\text{He}$  as a reliable cosmological baryometer. Their estimate of primordial baryon density is in excellent agreement with that

derived from the low abundance of deuterium measured in the absorption spectra of high-redshift quasars<sup>6</sup>, and from the high content of  $^4\text{He}$  in H II regions in metal-poor dwarf galaxies<sup>7</sup> (Fig. 1). It implies that  $^7\text{Li}$  has been depleted by about a factor of two in the oldest stars in our Galaxy. The estimate also agrees with an independent determination based on measurements of temperature anisotropy in the cosmic microwave background radiation by the MAXIMA and Boomerang experiments<sup>8</sup>.

That estimate, however, is low — only about 4% of the amount required to close the Universe, assuming a value for the Hubble constant of  $65 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . So the answer to the question we started with is that the Universe cannot be closed with baryonic matter only. But other evidence implies the existence of 'dark matter', or 'missing mass', the search for which now constitutes one of the most exciting areas of research at the interface between particle physics and cosmology. ■

Corinne Charbonnel is at the Laboratoire d'Astrophysique CNRS, Observatoire Midi-Pyrénées, 14 avenue Edouard Belin, F-31400 Toulouse, France.

e-mail: corinne@ast.obs-mip.fr

1. Yang, J., Turner, M. S., Schramm, D. N., Steigman, G. & Olive, K. A. *Astrophys. J.* **281**, 493–511 (1984).
2. Bania, T. M., Rood, R. T. & Balser, D. S. *Nature* **415**, 54–57 (2002).
3. Rood, R. T., Steigman, G. & Tinsley, B. M. *Astrophys. J.* **207**, L57 (1976).
4. Gloecker, G. & Geiss, J. *Space Sci. Rev.* **84**, 275–284 (1998).
5. Charbonnel, C. & do Nascimento, J. D. Jr *Astron. Astrophys.* **336**, 915–919 (1998).
6. Kirkman, D., Tytler, D., Burles, S., Lubin, D. & O'Meara, J. M. *Astrophys. J.* **529**, 655–660 (2000).
7. Izotov, Y. I. & Thuan, T. X. *Astrophys. J.* **500**, 188–216 (1998).
8. Stomp, R. *et al.* *Astrophys. J.* **561**, L7–L10 (2001).
9. Tytler, D. *et al.* *Physica Scripta* **T85**, 12–31 (2000).

## Neuroscience

# Sounds, signals and space maps

Catherine Carr

The auditory system transforms information from one frame of reference into another to create a map of space in the brain. The source of a visual signal that guides this transformation in barn owls has now been found.

When a barn owl hears a noise, it can pinpoint where the sound came from because its brain has a mental map of how the noises it hears fit into the space around it. Various auditory cues are used to construct this 'auditory space map', but it must also be guided by visual inputs, because — assuming reasonable eyesight — vision provides more reliable, topographically organized information. But what part of the brain transmits this information to the auditory map, and what form does the information take? On page 73 of this issue Hyde and Knudsen<sup>1</sup> provide an answer. They suggest that a signal from the optic tectum region of the barn owl's brain pro-

vides topographic, point-to-point instructions about the correct representation of auditory spatial cues.

For a hungry barn owl, it is vital to be able to use sounds to locate potential prey — hence the importance of the mental map of those sounds in space. This map is found in the midbrain, specifically the optic tectum. But it is initially produced nearby, in the inferior colliculus<sup>2</sup>, using auditory localization cues to tune individual neurons. Such cues include differences in the time it takes for sounds to reach each of the ears, and differences in the level of sounds that arrive at each ear. The map is then relayed to the overlying optic tectum, where it is aligned





### 100 YEARS AGO

Dr. Sven Hedin, the Swedish explorer, who recently arrived at Ladakh from Central Asia, has sent a telegram to King Oscar announcing that he has made an extremely important journey through all Tibet, disguised as a pilgrim, with two followers. On approaching Lhasa they were recognised and captured, but were well treated by order of the Dalai Lama. A second attempt was opposed by 500 Tibetan soldiers. Dr. Hedin's collections were lost, with almost the whole caravan, but his notes were saved.

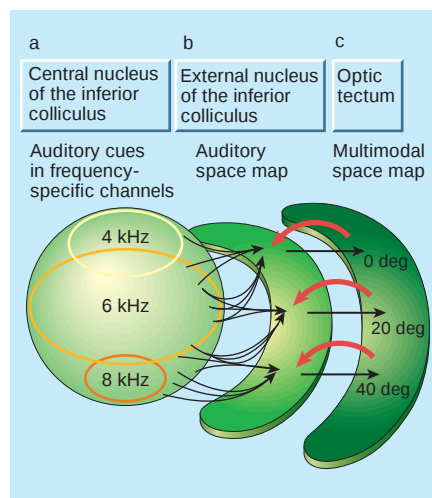
From observations described by Prof. Geitel in the *Physikalische Zeitschrift*, iii, 4, it appears that atmospheric air is itself capable of inducing radio-activity. When a mass of air remains shut up for a long time in a cellar or a cave, Prof. Geitel finds that its electric conductivity increases to a maximum. There are three hypotheses possible, namely, that the exposed substances were themselves radio-active, that traces of radio-active substances were present in the neighbourhood, or that the air itself is the origin of the radio-activity... Prof. Geitel favours the view that the third is the most likely hypothesis.

Many animals have popular names which have been derived from their cries. Prof. T. D. A. Cockerell writes to suggest that this is also the case with the donkey, the "don" representing the inspiratory and "key" the expiratory sound. Most dictionaries describe the word, which is of comparatively recent origin, as signifying a little dun animal, from dun and the diminutive term — key, but the grounds upon which this derivation is based are not easy to find.

From *Nature* 2 January 1902.

### 50 YEARS AGO

On September 29, 1951, Dr. Seth B. Nicholson, while examining a plate exposed earlier that night, discovered an object which he thought was a new satellite of Jupiter. No announcement was made until the object had been photographed later on three occasions with the 60-in. telescope at Mount Wilson and also with the 100-in., with which Dr. Nicholson had made the first photograph... The magnitude of the new satellite is 18.3, but this is only a provisional figure; its diameter has been estimated to be about 15 miles — the diameter of satellite X. Dr. Nicholson now ranks with Galileo as the only astronomer who has discovered four of Jupiter's satellites. From *Nature* 5 January 1952.



**Figure 1 Translating auditory cues into a map of 'auditory space' in barn owls.** a, Sounds are first transmitted to the brain's central nucleus of the inferior colliculus, where they are grouped by various criteria such as frequency. b, The resulting neural signals are then transmitted into the external nucleus of the inferior colliculus, where they are organized into an arrangement that represents the location of the sounds in space. c, The auditory map is also modified by visual signals (larger arrows), which Hyde and Knudsen<sup>1</sup> now show to be organized topographically and to come from the optic tectum. The resulting auditory map is then transmitted to the optic tectum and aligned with a visual map, producing a multimodal map of space. The values in degrees refer to the space around the animal.

with a visual map and used to generate orientating movements — when the owl focuses its eyes on a mouse, for example.

Most brain maps are 'projection' maps, in which neurons in the retina or skin, say, project to the relevant brain area in such a way that the spatial arrangement of neurons is the same<sup>3</sup>. In contrast, the auditory space map in the inferior colliculus does not reflect the organization of the auditory nerves, but is actively computed when information from the two ears is combined in the brain (Fig. 1). The translation of auditory cues into a topographic representation of space is clearly a complex transformation, even for a static auditory 'scene'. And the transformation must also be adaptable — able to respond to changes in the environment.

Auditory space maps can be generated without visual input, but their precision and topography depend on visual experience. So, for example, owls raised as if they were blind end up with abnormal, or even partially inverted<sup>4</sup>, auditory maps. And vision does not simply guide map formation. It also wins out when there is a mismatch between visual and auditory cues, as Knudsen and Brainard<sup>5</sup> showed in a classic experiment in which they raised young owls with prisms mounted in spectacle frames in front of the eyes. The prisms provided erroneous visual signals that caused a perfectly good auditory map to shift.

Because visual experience 'calibrates' the auditory space map, there must be an instructive signal from the visual system. Two years ago, Hyde and Knudsen<sup>6</sup> and Luksch *et al.*<sup>7</sup> proposed that the source of this signal is the optic tectum. It seems that neurons in this region project to the auditory space map in the inferior colliculus, in such a way as to connect points in the visual and auditory maps that represent the same part of space. These projections form before the owls hatch. The projecting neurons have some extensions that lead into the superficial tectal layers (which receive input directly from the retina) and some that extend into the deep tectal layers (which receive auditory

input from the auditory space map and visual input from the forebrain). From these anatomical data, it seemed likely that the optic tectum could provide the visually based signal that guides the auditory map. But this was not certain.

There are two types of visually based instructive signals that could guide adaptive adjustments of the auditory space map. These are a 'global value signal', derived from a visual assessment of how accurately the brain relates auditory and visual stimuli to locate the source of a sound; and a simple spatial template, derived directly or indirectly from a topographic map of visual space<sup>8</sup>. Either signal could account for the effects of the prisms on the auditory space map. But they would have different neural substrates: a global value signal would require a diffuse projection to the auditory space map; and a spatial template would require a point-to-point projection. So the earlier results<sup>6,7</sup> support the idea of a spatial template.

Hyde and Knudsen<sup>1</sup> now confirm this, in work that not only reveals the source of the instructive signal but also distinguishes between the two types of possible signals. They show that the nervous system indeed uses point-to-point neuronal projections from the tectum as a spatial template to guide the topography of the auditory space map.

First, the authors placed prisms that displaced the visual field on barn owls for six weeks. The owls' auditory space map gradually shifted to match their displaced visual map. The authors then damaged a small region in the owls' tectum, and removed the prisms so that the visual (and hence auditory) map could adapt. They found that most of the auditory space map could still change in response to the new visual experience. But the portion of the map that corresponded to the damaged part of the visual map was stuck, and could not adapt. These results show definitively that the optic tectum provides a visual signal to guide the transformation of auditory cues into an auditory space map. Moreover, the signal has a topographic

basis. How exactly this signal instructs the auditory map is still unclear, however.

In terms of understanding the fundamentals of learning and neuronal plasticity (the ability to be modified), the real excitement now lies in what can be done with this signal. If researchers can learn how to activate it, then they can begin to unravel the mechanisms by which one set of neurons provides a teaching signal that drives and controls plasticity in another. ■

Catherine Carr is the Department of Biology,

University of Maryland, College Park, Maryland 20742-4415, USA.

e-mail: cc117@umail.umd.edu

1. Hyde, P. S. & Knudsen, E. I. *Nature* **415**, 73–76 (2002).
2. Knudsen, E. I. & Konishi, M. *Science* **200**, 795–797 (1978).
3. Kaas, J. H. & Hackett, T. A. *J. Comp. Neurol.* **421**, 143–145 (2000).
4. Knudsen, E. I., Esterly, S. D. & du Lac, S. J. *Neurosci.* **11**, 1727–1747 (1991).
5. Knudsen, E. I. & Brainard, M. S. *Science* **253**, 85–87 (1991).
6. Hyde, P. S. & Knudsen, E. I. *J. Comp. Neurol.* **421**, 146–160 (2000).
7. Luksch, H., Gauger, B. & Wagner, H. J. *Neurosci.* **20**, RC70 (2000).
8. Rucci, M., Tononi, G. & Edelman, G. M. *J. Neurosci.* **17**, 334–352 (1997).

## Astronomy

# Magnetic bubbles in space

Ellen G. Zweibel

The origin of magnetic fields found in galaxies and galaxy clusters is unknown. Both models and observations suggest that extinct radio galaxies could be responsible.

**M**agnetic fields are found in galaxies and clusters of galaxies throughout the Universe. There is even some evidence for magnetized plasma (ionized gases) bridging the space between two galaxy clusters. But the origin of these magnetic fields and their influence on the intergalactic environment remains largely a mystery<sup>1</sup>. Theories of how these fields are generated follow either the top-down model, in which magnetization occurred on a cosmic scale when the Universe was much younger, or the bottom-up idea, in which magnetic fields are generated on small scales and then merge into larger systems. Two papers in the *Astrophysical Journal* by Furlanetto and Loeb<sup>2</sup> and Kronberg *et al.*<sup>3</sup> propose that the fossil remnants of extragalactic radio sources fill a sizeable fraction of the intergalactic medium, supporting the bottom-up approach.

Strong radio sources, such as quasars, generate their radio emission by synchrotron radiation from electrons moving at relativistic speeds (close to the speed of light) in a magnetic field. So where there is strong radio emission there are almost always magnetic fields. The idea of radio galaxies polluting the intergalactic medium with magnetic fields is not entirely new<sup>4</sup>, but better theory<sup>2</sup> and observations<sup>3</sup> now allow for increasingly detailed calculations of their contribution to the magnetization of intergalactic space. At these large scales, such orphan magnetic fields could have a substantial influence on the formation and evolution of galaxies.

Quasars are the most luminous objects we know at optical, and in some cases radio, wavelengths. They are extremely distant active galactic nuclei that formed when the Universe was very young. According to the standard model, quasars shine so brightly because energy is efficiently radiated by material accreting from an orbiting disk

onto a massive black hole in the nucleus<sup>5</sup>. Roughly 10 to 20% of quasars eject plasma and magnetic fields into their surroundings in the form of powerful jets and winds moving at relativistic speeds. The jets feed into diffuse 'lobes', which are clouds of radio-emitting material often found on either side of the compact core, forming a double-lobed structure. The radio lobes can be several million light years in radius, so together they span a region perhaps 100 times larger than the parent galaxy (Fig. 1). The magnetic fields are believed to be generated within the black-hole accretion disks<sup>6,7</sup>, which are about the size of the Solar System (although the fields may be amplified further within the jets).

What happens to the radio lobes when the quasars die, which is thought to occur after roughly 10<sup>8</sup> years? Furlanetto and Loeb<sup>2</sup> and Kronberg *et al.*<sup>3</sup> argue that they continue to expand, acting as magnetized bubbles that sweep up intergalactic material into shells. Although there are many uncertainties, such as the pressure and density of the surrounding intergalactic medium, the model developed by Furlanetto and Loeb<sup>2</sup> suggests that up to 10–20% of intergalactic space could be filled with these magnetized fossils. The original synchrotron electrons lose their energy as the bubble expands, causing the radio emission to fade, but shock waves from shell formation could accelerate them to high energies. Or the bubbles could lurk in obscurity until they are recompressed and re-energized by the sort of large-scale collisions associated with the formation of galaxy clusters.

Although the broad outlines of bubble expansion now seem to be understood, their interaction with the intergalactic environment is less certain. To address this issue, Kronberg *et al.*<sup>3</sup> analysed almost 100 extra-

galactic radio sources, some of which are found in low-density intergalactic space, whereas others are located near the cores of dense galaxy clusters — two extremes of the intergalactic environment. The authors use images of the luminosity and volume of their radio lobes to calculate the amount of magnetic energy stored in each source. They find that the magnetized radio lobes in the cluster region have to work much harder to expand, and so contain much less magnetic energy than the isolated radio sources. Nonetheless, the giant radio lobes in less dense areas of space could be expected to expand and magnetize a large fraction of the intergalactic medium.

The final character of these magnetized fossils and their impact on their intergalactic neighbours remains unknown. The magnetic fields could affect galaxy formation and evolution if they are strong enough to exert significant dynamical forces. The orbits of cosmic rays (highly energetic particles that can travel intergalactic distances) are also affected by magnetic fields: a cosmic ray encountering a large magnetic bubble would be swung around by it, changing its direction of motion. If there were enough bubbles, the cosmic rays would be dispersed rather than just heading out in straight lines. This is important for theories of the origin of the highest-energy cosmic rays, which appear to have travelled enormous distances. However, the source of these particles could be much closer to our own galaxy if magnetic

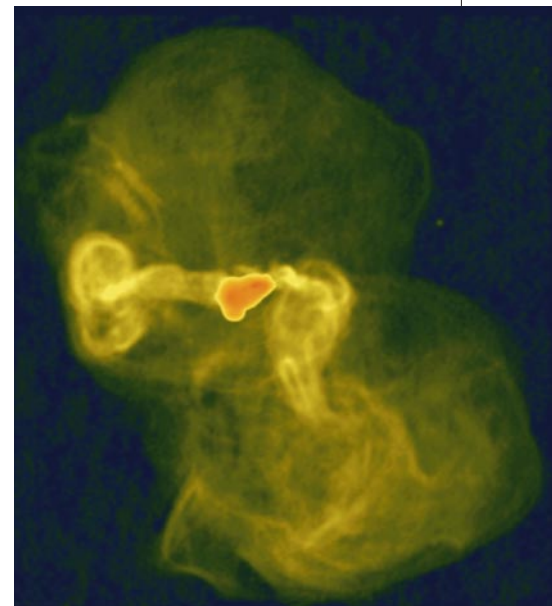


Figure 1 Halo around an active radio galaxy. This giant elliptical galaxy (M87) is one of the strongest radio sources in the sky. The radio halo can be seen as a green/yellow bubble against the dark blue X-ray-emitting background. The magnetized bubble is about 80,000 light years across. The black hole at the centre of the inner structure (red) is connected to the surrounding halo by jets of plasma and magnetic fields.

E. N. OWEN, J. A. ELIEK, N. E. KASSIM, NRAO



bubbles had dispersed them, because the linear distance they travel would be much smaller.

But there is one major problem with any bottom-up theory of magnetic-field formation, which has to do with the existence of positive and negative fields. Because field directions in different fossils vary randomly, whereas positive and negative fields within each magnetized bubble presumably cancel each other out, opposing fields could be dragged into proximity and annihilate one another, leaving behind almost nothing. But the overall rate of field destruction is difficult to estimate — the true picture could range from the survival of isolated bubbles, to partial mergers between magnetic systems in which the fields wander over vast regions of

space, to the nearly field-free situation in which total cancellation has occurred. In any event, active galaxies continue to spew out magnetic fields into the intergalactic medium, and astronomers continue to extend their observations and models, so one day we may be able to confirm the bottom-up picture. ■

Ellen G. Zweibel is at JILA, University of Colorado, Boulder, Colorado 80309, USA.

e-mail: zweibel@lepton.colorado.edu

1. Zweibel, E. G. & Heiles, C. *Nature* **385**, 131–136 (1997).
2. Furlanetto, S. R. & Loeb, A. *Astrophys. J.* **556**, 619–634 (2001).
3. Kronberg, P. P., Dufton, Q. W., Li, H. & Colgate, S. A. *Astrophys. J.* **560**, 178–186 (2001).
4. Rees, M. J. & Setti, G. *Nature* **219**, 127–128 (1968).
5. Begelman, M. C., Blandford, R. D. & Rees, M. J. *Rev. Mod. Phys.* **56**, 255–351 (1984).
6. Colgate, S. A. & Li, H. *Astrophys. Space Sci.* **264**, 357 (1999).
7. Hawley, J. *Astrophys. J.* **528**, 462–479 (2000).

## Microbiology

# Group effort in toxin synthesis

Gary M. Dunny

It is increasingly evident that bacterial cells cooperate for many purposes. New results show that the bacterium *Enterococcus* uses cell–cell signalling to coordinate toxin production.

Hospital patients with diseases that impair their natural defences are susceptible to opportunistic infections. This is a serious medical problem. Opportunistic bacteria, such as *Enterococcus faecalis*, are often resistant to antimicrobial drugs. And because these bacteria do not generally cause disease in healthy people, little is known about why they infect and cause disease in hospital patients. On page 84 of this issue<sup>1</sup>, Haas *et al.* reveal a previously unknown way in which enterococcal cells communicate with one another to coordinate toxin production. The results shed light on bacterial cell–cell signalling mechanisms, and may also provide insights into enterococcal pathogenicity and open up new approaches to drug discovery.

In the past, microbial life was largely viewed as individual organisms single-mindedly pursuing growth by cell division. But it is now clear that many microbial activities — including genetic exchange, growth in biofilms, expression of virulence factors (toxins) and antibiotic production — result from well-coordinated multicellular processes controlled by cell–cell signalling<sup>2</sup>. This applies to both of the two main groups of bacteria, Gram-negative and Gram-positive (this division is made on the basis of how their cell walls absorb a specific stain).

In Gram-negative bacteria — *Pseudomonas*, *Agrobacterium* and *Vibrio*, for instance — the best-studied signal molecules are acylhomoserine lactones that diffuse into the bacterial cell and interact with a gene-transcription factor to increase synthesis of

messenger RNA from the relevant target genes. Lactone synthesis occurs at a low level in bacterial cultures, but when bacteria are growing exponentially lactone concentration increases with cell density. When a sufficient concentration is reached, the target genes (often including the synthetase gene for acylhomoserine lactone) are transcribed at high levels, causing a dramatic 'auto-induction' of the same activity throughout the population. Such cell-density-dependent microbial behaviour, mediated by signal molecules, is called 'quorum sensing'<sup>3</sup>.

Gram-positive bacteria, such as streptococci, bacilli, staphylococci and enterococci, also carry out quorum sensing, as well as related forms of cell–cell communication between different cell types (for instance, DNA transfer between donor and recipient cells). In these organisms, however, the signal molecules are peptides of about 5–20 amino acids in length<sup>4</sup>. To date, two types of signalling pathway have been found.

In most cases, processing of the peptide signal involves a membrane sensor protein, whose interaction with the signal molecule results in autophosphorylation of a histidine residue in the protein (Fig. 1a, overleaf). The phosphorylated sensor (one component) communicates with a response-regulator protein (a second component) in the cytoplasm, transferring the phosphate to the regulator and converting it to an active transcription factor that interacts with target promoters to activate gene transcription. These two-component signal-transduction systems are members of a large family of

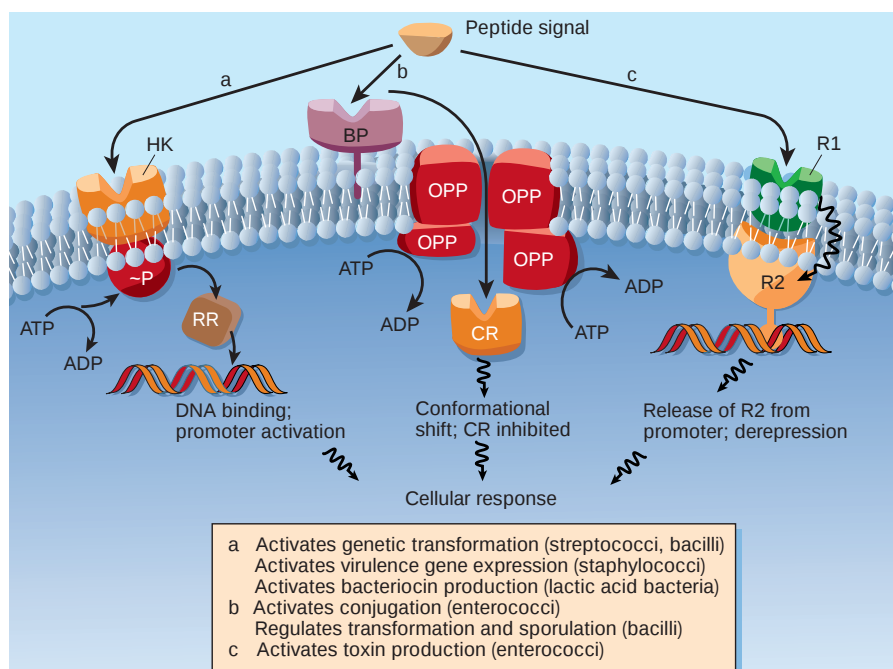
regulators that control bacterial responses to a variety of environmental signals<sup>5</sup>. In a second type of pathway, the peptide signal is imported into the responder cell, where its interaction with an intracellular receptor generates the response (Fig. 1b). The details are not known, but it seems that the imported peptide can act in various ways in the target cell<sup>4</sup>.

The results of Haas *et al.*<sup>1</sup> are based on analysis of a toxin called cytolyisin, which this group has studied for several years<sup>6–8</sup>. The toxin has two components, LL and LS. When released from the cell, these components act in concert to attack a variety of bacterial and mammalian target cells. But before they are released, both are processed by a variety of accessory components.

A cluster of at least half-a-dozen genes is required to produce cytolyisin and to protect the enterococcal cell against its effects. The cluster is generally encoded by accessory mobile genetic elements, in this case a plasmid called pAD1 (refs 9, 10). Interestingly, expression of the *pAD1* transfer genes is regulated by donor–recipient cell signalling mediated by the peptide cAD1 (ref. 8). Production of cytolyisin increases enterococcal virulence in several experimental systems, and cytolyisin is commonly, but not universally, found in enterococcal strains isolated from hospital patients<sup>11,12</sup>.

Haas *et al.*<sup>1</sup> analysed a pair of genes, *cytR1* and *cytR2*, that are located next to the gene cluster responsible for cytolyisin biosynthesis but are transcribed separately. Using a carefully constructed series of gene knockouts, and fusions between 'reporters' of gene transcription, they found that the *cytR1* and *cytR2* gene products repress expression of the biosynthetic genes. During growth in culture, this repression is overcome in a density-dependent fashion by a secreted bacterial product, which the authors identified as the *cytS*-encoded component (LS) of the toxin.

Computer analysis<sup>1</sup> of the predicted amino-acid sequences of the *cytR1* and *cytR2* products was particularly illuminating. The first product is predicted to be a small transmembrane protein; the second, a cytoplasmic protein with a DNA-binding motif. Neither molecule has much resemblance to the proteins of the two-component family of signal transducers, or to any other proteins in the databases. The authors propose that, in the default state of the system, interaction between a cytoplasmic segment of R1 and R2 holds R2 in a repressing conformation and biosynthesis occurs only at a low level. As cell density reaches about 10<sup>7</sup> cells per millilitre, the concentration of LS in the medium becomes sufficient to cause induction of cytolyisin biosynthetic genes, presumably by direct interaction with an extracellular domain of R1 and a conformational shift which disrupts the R1–R2 interaction (Fig. 1c). These molecular interactions have



**Figure 1 Three mechanisms of cell-cell communication between bacteria.** The signalling molecules are peptides, 5–25 amino acids in length, but are different for each pathway. **a**, A two-component signal-transduction system, in which binding of the peptide to a histidine kinase (HK) receptor on the cell membrane results in autophosphorylation (~P) of the kinase. The phosphate is transferred to a response regulator (RR), which can then bind to target DNA and activate a gene promoter and transcription. **b**, A binding protein (BP) on the cell membrane takes the peptide signal from the environment and uses an oligopeptide permease (OPP) to import the peptide into the cell. Peptide binding to a cytoplasmic receptor (CR) causes a conformational shift and inhibits receptor function (some CRs negatively regulate transcription; others act as phosphatases, with phosphorylated regulatory proteins as substrates). **c**, The pathway described by Haas *et al.*<sup>1</sup>. In the absence of the peptide signal (in this case, the LS component of the secreted cytotoxin), two proteins, R1 and R2, appear to act in concert to repress the cytotoxin biosynthetic genes. Peptide-signal binding to R1 is predicted to disrupt the repressive R1–R2 conformation and result in increased gene transcription. Squiggly arrows indicate the effects of interactions; other arrows indicate direct molecular interactions. Bacterial activities controlled by each pathway are listed in the box.

not been demonstrated directly, but they are consistent with the data and can be tested experimentally.

These results will affect our view of cell-cell communication in bacteria, and how we deal with the *Enterococcus* as an opportunistic pathogen. Clearly, peptide-mediated cell-cell signalling in bacteria can occur by any of at least three distinct mechanisms (Fig. 1). A search of the nearly completed *E. faecalis* genome sequence<sup>13</sup> suggests that there may be a dozen or more two-component signal-transduction systems, some of which could mediate cell-cell communication. In other words, in *E. faecalis* alone, many additional processes could be regulated in this fashion.

The new results<sup>1</sup> could also help in developing drugs to combat enterococcal infections. In some infections, production of cytotoxin or other virulence factors (regulated, perhaps, by similar mechanisms) may be required to cause disease but may not be necessary for the organism's survival and growth in its normal habitat, the intestine. A drug that targeted the signal pathway to block virulence could prevent or cure infection. But because the bacteria would

not be directly killed, the drug would not exert a strong selective pressure for development of drug resistance — one of the banes of modern medicine.

Gary M. Dunny is in the Department of Microbiology, University of Minnesota Medical School, 428 Delaware Street, SE, Minneapolis, Minnesota 55455-0312, USA.

e-mail: gary-d@biosci.cbs.umn.edu

1. Haas, W., Shepard, B. D. & Gilmore, M. S. *Nature* **415**, 84–87 (2002).
2. Dunny, G. M. & Winans, S. C. *Cell-Cell Signaling in Bacteria* (Am. Soc. Microbiol., Washington, DC, 1999).
3. Fuqua, W. C., Winans, S. C. & Greenberg, E. P. *J. Bacteriol.* **176**, 269–275 (1994).
4. Dunny, G. M. & Leonard, B. A. *B. Annu. Rev. Microbiol.* **51**, 527–564 (1997).
5. Hoch, J. A. & Silhavy, T. J. (eds) *Two-Component Signal Transduction* (Am. Soc. Microbiol., Washington, DC, 1995).
6. Ike, Y., Clewell, D. B., Segarra, R. A. & Gilmore, M. S. *J. Bacteriol.* **172**, 155–163 (1990).
7. Gilmore, M. S., Segarra, R. A. & Booth, M. C. *Infect. Immunol.* **58**, 3914–3923 (1990).
8. Segarra, R. A., Booth, M. C., Morales, D. A., Huycke, M. M. & Gilmore, M. S. *Infect. Immunol.* **59**, 1239–1246 (1991).
9. Clewell, D. B. *et al.* *J. Bacteriol.* **152**, 1220–1230 (1982).
10. Gilmore, M. S. *et al.* *J. Bacteriol.* **176**, 7335–7344 (1994).
11. Huycke, M. M. & Gilmore, M. S. *Plasmid* **34**, 152–156 (1995).
12. Coque, T. M., Patterson, J. E., Steckelberg, J. M. & Murray, B. E. *J. Infect. Dis.* **171**, 1223–1229 (1995).
13. <http://www.tigr.org>

## Daedalus

### Atmospheric charge

Lightning, says Daedalus, is universal. Not only are there about a thousand thunderstorms going on around the Earth at any one time, but lightning has been observed in the atmospheres of other planets, with different chemistry from that of Earth. The complex organic molecules built up by earthly lightning, precursors of life itself, should therefore exist on other planets. They may have given life a start there too.

Daedalus can think of two bases for lightning. First, phase changes, like that of water going to ice, must extrude charge. No solution or melt can have the same electron affinity as the solid, so a solidifying droplet must push charge ahead of the solid front. Second, any charged gas must expand, by self-repulsion. A big enough voltage (150 megavolts for air) would give a vacuum; smaller charges, of just a few megavolts, would merely provide strong lift. So solidification of any sort, coupled with charge sharing to an atmosphere, must give rise to convection. Couple this with a feedback mechanism, such as the return of charge in falling raindrops that partly freeze, and you soon build up enough charge separation for lightning.

To test these ideas, DREADCO physicists are building a large-scale lightning machine. They are starting with a hollow cooling tower. At the top, water will be sprayed down it from perforated pipes. It will lose heat and should partly freeze. The charge thus generated will be expelled into the remaining liquid. This should pass on charge to the surrounding air, which will expand and rise in its turn. At the top of the tower, copper electrodes will capture charge from the rising air and return it to the falling drops. In this way the tower will build a charge separation. It may also be done with brine. Brine is a conductor, but its freezing expels solid, insulating salt. Again, the physicists will look for the accumulation of charge. Any surplus high-voltage electricity from the tower will be bled from its top for use.

Sadly, Daedalus's tower may not generate much power. Even a full-scale thunderstorm only creates some 20 gigawatts. But he hopes to identify the electrostatic processes that are important in lightning, on Earth or on other planets. The expansion of charged gas, and its resulting speed of uplift, are well worth knowing; so is the feedback of charge by falling drops. Both may help us to understand the biochemistry of other planets.

David Jones



# Seeing through the face of deception

Thermal imaging offers a promising hands-off approach to mass security screening.

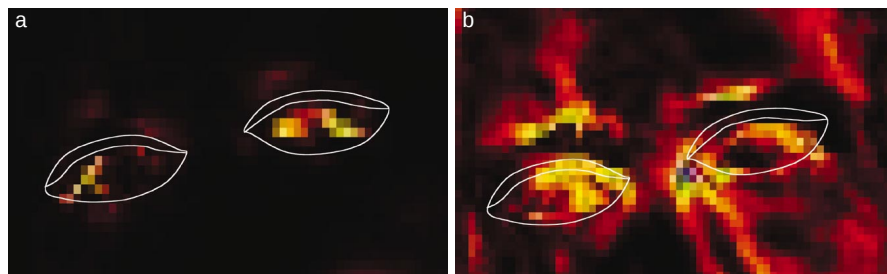
**W**e have developed a high-definition thermal-imaging technique that can detect attempted deceit by recording the thermal patterns from people's faces. This technique has an accuracy comparable to that of polygraph examination by experts and has potential for application in remote and rapid security screening, without the need for skilled staff or physical contact.

There is an urgent need to devise technologies that can be used for automated, high-throughput screening to identify individuals intending to perform acts of terrorism. At present, practicalities dictate that we rely on subjective assessment of responses to brief questions such as "Did you pack your own bags?" and "Why are you entering this facility?"

Although polygraph examinations, which have high precision when applied by experts<sup>1</sup>, are good at identifying liars, they are impracticable for mass screening because skilled operators are needed, subjects have to be attached to instrumentation for several minutes, data analysis is time-consuming and the interpretation of data is delayed.

We explored the possibility of using high-definition thermal imaging of the face for detecting deceit<sup>2</sup> because it enables rapid automated analysis of changes in regional facial blood flow to be quantified<sup>3,4</sup>. We have shown previously<sup>2</sup> that auditory startling is associated with a specific facial 'thermal signature', in which there is instantaneous warming around the eyes — probably as part of a fright/flight response mediated by the sympathetic nervous system<sup>5,6</sup>. Although the psychophysiology of startling differs from volitional deception, the nonspecificity of this facial thermal signature is reminiscent of the nonspecific variables monitored during a polygraph (respiration, pulse, relative blood pressure and electrodermal response). Were this thermal signature to accompany lying, independently of startling, it could be used for instantaneous lie detection without the subject even being aware of the test.

We therefore asked volunteers to commit a mock crime and then testify to their innocence under experimental conditions at the US Department of Defense Polygraph Institute (DoDPI; <http://www.dodpi.army.mil>)<sup>7</sup>. Twenty individuals were randomly assigned to stab a mannequin, rob it of \$20 and then assert their innocence of the 'crime'. Control subjects had no knowledge of the crime or of the crime scene. The thermal imaging system correctly categorized 83% of these subjects (Fig. 1); three-quarters (6 of 8) of the guilty individuals were correctly



**Figure 1** Periorbital, high-resolution thermal images of the face of a 'guilty' subject. Images were obtained before (a) and after (b) lying in reply to the question "Did you steal the \$20?" Images were obtained at 30 frames per second with a cooled thermal camera with a thermal sensitivity of 0.025 °C. The camera was calibrated daily to  $T_{min} = 29.00$  °C (black) and  $T_{max} = 38.00$  °C (cyan) with an external black body; red, orange and yellow represent progressively warmer temperatures in between. White lines indicate eye contours.

identified as guilty and 90% (11 of 12) of the innocent individuals were correctly categorized as innocent. Traditional polygraphs, performed by experts at DoDPI on the same subjects, correctly categorized 70% of the subjects: 6 of 8 subjects were correctly identified as guilty and 8 of 12 were correctly identified as innocent. Under these experimental conditions, the accuracy of the thermal imaging system was comparable to that of the traditional polygraph.

High-definition thermal imaging of the face is therefore a promising technology that should allow psychological responses to be detected and analysed rapidly and without physical contact, in the absence of trained

staff and in a variety of different situations. **Ioannis Pavlidis\***, **Norman L. Eberhardt†**, **James A. Levine†**

\*Honeywell Laboratories, 3660 Technology Drive, Minneapolis, Minnesota 55418, USA

†Mayo Clinic, Endocrine Research Unit, Department of Medicine, Minnesota 55905, USA  
e-mail: [levine.james@mayo.edu](mailto:levine.james@mayo.edu)

1. Kircher, J. C., Horowitz, S. W. & Raskin, D. C. *Law Hum. Behav.* **12**, 79–90 (1988).
2. Levine, J. A., Pavlidis, I. & Cooper, M. *Lancet* **357**, 1757 (2001).
3. Gratt, B. M. & Sickles, E. A. *J. Orofac. Pain* **9**, 255–265 (1995).
4. Barrett, A. H. & Myers, P. C. *Science* **190**, 669–671 (1975).
5. Fendt, M. & Fanselow, M. S. *Neurosci. Biobehav. Rev.* **23**, 743–760 (1999).
6. Otake, K., Ruggiero, D. A. & Nakamura, Y. *Brain Res.* **697**, 17–26 (1995).
7. Holden, C. *Science* **291**, 967 (2001).

## Satellite tagging

### Expanded niche for white sharks

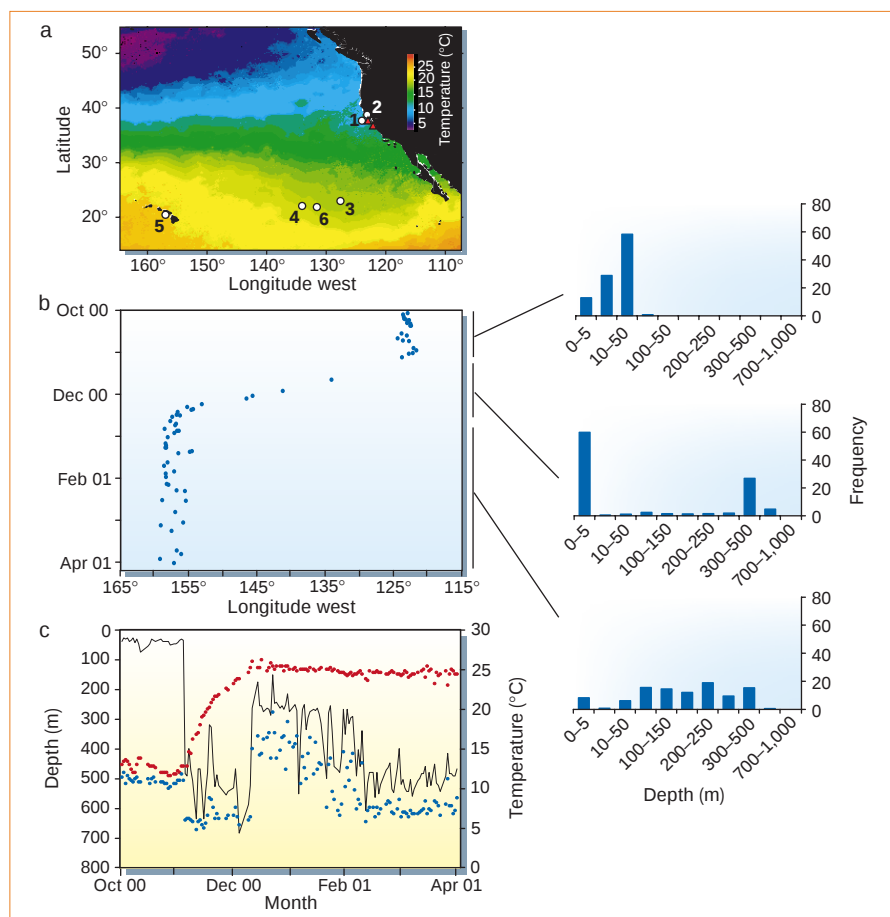
**U**ntil the advent of electronic tagging technology<sup>1–4</sup>, the inherent difficulty of studying swift and powerful marine animals made ecological information about sharks of the family Lamnidae<sup>5,6</sup> difficult to obtain. Here we report the tracking of movements of white sharks by using pop-up satellite archival tags, which reveal that their migratory movements, depth and ambient thermal ranges are wider than was previously thought.

White sharks (*Carcharodon carcharias*) are globally distributed, and have been reported to inhabit primarily continental-shelf waters in temperate seas<sup>6</sup>. Most tracking studies, however, have been limited to seasonal investigations around coastal pinniped colonies<sup>7–9</sup>. We have extended these over much wider ranges by retrieving data from pop-up satellite archival tags applied to the dorsal musculature of six adult white sharks (3.7–5.0 m in length) caught off the

coast of central California. The tags collected pressure, temperature and light-level data at 2-min intervals over a cumulative 650 days (see supplementary information). Light-level data were used to estimate local midnight or noon for longitude calculations<sup>10,11</sup>. At a pre-programmed date, the tags detached from the fish and transmitted a summary of stored data through the Argos satellite system.

We tagged six sharks in 1999–2000 and tracked them for periods ranging from 0.5 to 6 months (Fig. 1a). All sharks underwent a near-shore phase immediately after tagging. Diving patterns and ambient-temperature preferences during the coastal-residence period were similar for all sharks, who spent most of their time between the surface and a depth of 30 m, with the deepest dives reaching 75 m (Fig. 1b, c). During this period, the sharks experienced a narrow ambient water-temperature range of 10–14 °C.

Four sharks, which we tracked for 4–6 months, then moved offshore, where they remained exclusively pelagic. One individual (shark 5) travelled 3,800 km to waters off the western coast of the Hawaiian island of



**Figure 1** Movements, diving and temperature preference of white sharks. **a**, Deployment (red triangles) and end-point locations (white circles) for sharks tagged with pop-up satellite archival tags. Deployment dates are given first, followed by pop-off dates. 1, 19 October 1999; 2, 30 October 1999; 25 November 1999. 3, 16 October 2000; 19 February 2001. 4, 10 December 2000; 9 April 2001. 5, 16 October 2000; 16 April 2001. 6, 5 November 2000; 8 May 2001. Sea-surface-temperature image is a weekly composite for 21–28 February 2001. **b**, Longitude and depth distribution of shark 5 over the course of its 182-day tracking period. **c**, Data for shark 5 over the course of the tracking period: black line, maximum daily depth; red points, sea-surface temperature; blue points, minimum daily temperatures. Coastal residence is indicated by shallow maximum depths (which correspond to the shark's position over a continental shelf), low sea-surface temperature and narrow ambient temperature range.

Kahoolawe; three others (sharks 3, 4 and 6) moved to a region of the subtropical eastern Pacific (Fig. 1a). All four sharks showed a period of bimodal preference for depths of 0–5 m and 300–500 m, spending up to 90% of the day in these depth ranges and little time at intermediate depths (Fig. 1b shows representative data for shark 5). As the sharks moved southwest, they increased their maximum diving activity and experienced a broader range of ambient temperatures. Sea-surface temperatures rose to 20–26 °C, and the minimum temperatures at maximum depths (650–680 m) dropped to 4.8 °C (Fig. 1c), suggesting that white sharks can tolerate a broad temperature range.

The shark that travelled to Hawaii crossed 32° of longitude in 40 days at a minimum velocity of 71 km per day (Fig. 1b). Although sightings of white sharks in Hawaiian waters are rare<sup>12</sup>, this individual remained in the vicinity for almost 4 months, primarily staying between the surface and 300 m throughout this period (Fig. 1b).

These data provide the most extensive

record so far of the ecological niche of white sharks. Our results indicate that their range is more pelagic than was previously thought, comprising an inshore continental-shelf phase as well as extensive oceanic travel. The offshore phase lasted for at least 5 months, suggesting that it is an important period in the life history of white sharks in the North Pacific. It is unclear whether these offshore movements, which include extensive deep dives, represent feeding or breeding migrations. Increased tracking using electronic tagging should provide more data about the movement patterns, habitat usage and potential fishery interactions of white sharks, as well as critical information needed for the conservation of this species.

**Andre M. Boustany\*, Scott F. Davis†, Peter Pyle‡, Scot D. Anderson‡, Burney J. Le Boeuff, Barbara A. Block\***

\*Tuna Research and Conservation Center, Stanford University and Monterey Bay Aquarium, Hopkins Marine Station, Pacific Grove, California 93950, USA  
e-mail: bblock@stanford.edu

†Institute of Marine Sciences, University of California at Santa Cruz, Santa Cruz, California 95064, USA

‡Point Reyes Bird Observatory, Stinson Beach, California 94970, USA

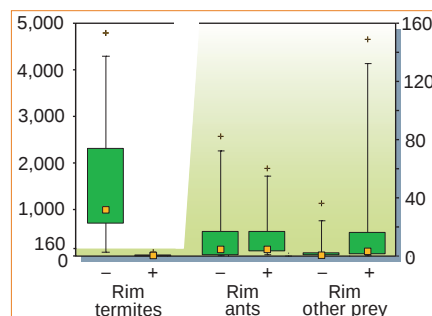
1. Metcalfe, J. D. & Arnold, G. P. *Nature* **387**, 665–666 (1997).
2. Block, B. A., Dewar, H., Farwell, C. & Prince, E. D. *Proc. Natl Acad. Sci. USA* **95**, 9384–9389 (1998).
3. Block, B. A. et al. *Science* **293**, 1310–1314 (2001).
4. Eckert, S. A. & Stewart, B. *Environ. Biol. Fish.* **60**, 299–308 (2000).
5. Carey, F. G. & Teal, J. M. *Comp. Biochem. Physiol.* **28**, 199–204 (1969).
6. Compagno, L. J. V. *FAO Fish. Synop.* **125**, 1–249 (1984).
7. Klimley, A. P. *Mem. South. Calif. Acad. Sci.* **9**, 15–40 (1985).
8. Goldman, K. J., Anderson, S. D., McCosker, J. E. & Klimley, A. P. in *Great White Sharks: The Biology of Carcharodon carcharias* (eds Klimley, A. P. & Ainley, D. G.) 111–120 (Academic, San Diego, 1996).
9. Klimley, A. P. et al. *Mar. Biol.* **138**, 617–636 (2001).
10. Hill, R. D. in *Elephant Seals: Population Ecology, Behavior, and Physiology* (eds LeBoeuf, B. J. & Laws, R. M.) 227–236 (Univ. California Press, Berkeley, 1994).
11. Welch, D. W. & Everson, J. P. *Can. J. Fish. Aquat. Sci.* **56**, 1317–1327 (1999).
12. Taylor, L. *Mem. South. Calif. Acad. Sci.* **9**, 41–48 (1985).

Supplementary information accompanies this communication on Nature's website (www.nature.com).

## Carnivorous plants

# Mass march of termites into the deadly trap

Carnivorous pitcher plants of the genus *Nepenthes* are not usually very selective about their prey, catching anything that is careless enough to walk on their slippery peristome, but *Nepenthes alboburginata* is an exception. We show here that this plant uses a fringe of edible white hairs to lure and then trap its prey, which consists exclusively of termites in enormous numbers. This singular feature accounts for the specialization of *N. albo-*



**Figure 1** Comparison of prey composition for pitchers with intact and with grazed-down rim hairs (box plot; rim condition: minus sign, grazed down; plus sign, intact). The prey groups 'ants' and 'other prey' (right) are presented on an extended scale. For statistical analysis, we used the non-parametric Mann–Whitney *U*-test. There is a significant difference in the number of termites ( $P > 0.02$ ), but no significant difference for the prey-group ants. The difference in the number of the group 'other prey' was significant ( $P > 0.02$ ) but in our opinion this was too heterogeneous to allow any conclusions to be drawn. Details are available from the authors. Plus signs, maximum values; hollow squares, medians; error bars, limits; green boxes, 25th to 75th percentiles.



*marginata* for one prey taxon, unique so far among carnivorous plants.

The main attractant for insect prey (ants are most common<sup>1-3</sup>) in most *Nepenthes* species is its nectar, which may be associated with a distinctive pitcher colour and smell<sup>1</sup>. But *N. albomarginata* T. LOBB ex LINDL works differently. If no termites are present, the harvest diversity looks much the same, except that the catch is even poorer than in neighbouring pitcher plants of other species: a few dozen ants, beetles or flies caught over the six-month lifespan of the pitcher. However, termites have frequently been found in *N. albomarginata* pitchers, a feature seldom, if ever, associated with other species<sup>3-5</sup>.

We studied *N. albomarginata* in Brunei and usually found thousands of termites trapped in a single pitcher (Fig. 1). All termites in one pitcher belonged to the same species and were in the same state of decomposition, suggesting that they were caught over a short period. We found three termite genera, all belonging to the subfamily Nasutitermitinae, with one genus (*Hospitalitermes*) predominating. The feeding by this termite on live fungal and algal tissue and its habit of foraging over ground in mass columns are extraordinary among termites.

*N. albomarginata* has a unique morphological feature, a rim of living white trichomes directly below the peristome (Fig. 2). This had not previously been associated with the plants' feeding habits until one of us (D.J.M.) noticed that the rim's hairs were missing from pitchers that had caught termites. A comparison of large numbers of

pitchers over a prolonged period confirmed the correlation. For several days, nothing would happen, then — after a single night — pitchers would fill with termites and their rim hairs would disappear.

To investigate this, we placed fresh intact pitchers, together with pitchers with their white rims removed, near to the head of foraging columns of the termite *Hospitalitermes bicolor*. When single leading workers came into direct contact with the white rim hairs, they turned back to the column and recruited their nestmates, which began grazing on the rim and forming food pellets from the white trichomes (Fig. 2). While doing this, the termites fell into the pitchers in their masses, workers and accompanying soldiers alike. Once in the pitcher, they were unable to grip the peristome to climb out. Observation of one grazing *H. bicolor* column revealed that up to 22 individuals per minute were falling into the pitchers, but the capture rate could easily exceed this for denser columns. Outside the pitcher traps, the termites at the column head displayed typical foraging and recruitment behaviour<sup>6,7</sup>. After about an hour, the hairs were all gone and the pitcher was evidently no longer attractive to termites (and was filled with termites trying to escape).

We do not know how the trichomes lure termites onto the plant. No long-range olfactory attraction could be detected during our experiments; all contacts seemed to happen by chance, with termites often missing pitchers less than 1 cm away from them.

Prey specialization has been proposed for some *Nepenthes* species<sup>3</sup> but, to our knowledge, *N. albomarginata* is the first example of a carnivorous plant in which this has been confirmed and functionally described, as well as being the only species to offer its tissue as a bait.

**Marlis A. Merbach\*, Dennis J. Merbach\*, Ulrich Maschwitz\*, Webber E. Booth†, Brigitte Fiala‡, Georg Zizka§**

\*Fachbereich Biologie, Zoologisches Institut, Johann Wolfgang Goethe-Universität, Frankfurt am Main, Postfach 111932, 60054 Frankfurt, Germany  
e-mail: merbach@zoology.uni-frankfurt.de

†Biology Department, Universiti Brunei Darussalam, Brunei Darussalam, Borneo

‡Universität Würzburg, Zoologie III, Biozentrum, Am Hubland, 97074 Würzburg, Germany

§Botanik/Paläobotanik, Johann Wolfgang Goethe-Universität und Forschungsinstitut Senckenberg, 60325 Frankfurt am Main, Germany

1. Moran, J. A. *J. Ecol.* **84**, 515–525 (1996).
2. Moran, J. A., Booth, W. E. & Charles, J. K. *Ann. Bot.* **83**, 521–528 (1999).
3. Kato, M., Hotta, M., Tamin, R. & Itino, T. *Trop. Zool.* **6**, 11–25 (1993).
4. Clarke, C. M. *Nepenthes of Borneo* (Natural History Publications (Borneo), Kota Kinabalu, Sabah, Malaysia, 1997).
5. Adam-Jumaat, H. *Pertanika J. Trop. Agric. Sci.* **20**, 121–134 (1997).
6. Jones, D. T. & Gathorne-Hardy, F. *Insectes Soc.* **42**, 359–369 (1995).
7. Miura, T. & Matsumoto, T. *Insectes Soc.* **44**, 267–275 (1998).

Competing financial interests: declared none.

COMMUNICATIONS ARISING

Geoarchaeology

## Did Nile flooding sink two ancient cities?

The discovery of the two cities of Herakleion and East Canopus under the waters of the Bay of Abu Qir (east of Alexandria, Egypt) stirred worldwide attention when it was first announced in the summer of 2000. Their disappearance some 1,250 years ago has been ascribed by Stanley, Goddio and Schnepf<sup>1</sup> to a strong Nile flood that caused riverbank failure and the destruction of the two cities, rather than to the action of earthquakes, as was first proposed when the ruins were discovered<sup>2,3</sup>. But I believe that this interpretation is flawed, because no flood could have reached the Abu Qir Bay at the time of the disappearance of the two cities, as the Canopic branch of the Nile, along whose banks they were situated, had dried to a trickle more than 200 years earlier.

This consideration seems to have been overlooked by Stanley *et al.*, who assume that the Canopic remained active until late in the first millennium. Toussoun, who is cited as the main reference for that assumption, is clear on this point: on pages 195–196 of his work<sup>4</sup>, he gives evidence that the Canopic silted up gradually during the fifth century and that it stopped flowing into the bay long before the Arab armies reached Alexandria; he found no mention of it in any of the accounts of the Arab historians.

The radiocarbon ages given by Stanley *et al.* for the mud “immediately beneath the ruins” confirm this view: none is younger than AD 322. I know of no mention of this branch of the delta past the fifth century AD, neither is there any evidence of it on the maps prepared by Napoleon's savants in the eighteenth century.

Even if we accept the authors' assumption that the Canopic was still active at the time of the disappearance of Eastern Canopus and Herakleion, it is unlikely that the destruction of these two cities would have been brought about so suddenly. There is no mention of such a cataclysmic event in any of the texts written around this period, nor of the destruction of any other city along the length of the Nile.

The flood in AD 741/742 was not likely to have been dangerous or destructive. For those who know the Nile and the history of its use, floods associated with higher water levels than normal were welcome events. Before the system of perennial irrigation was introduced in the nineteenth century, the lowlands of the flood plain of the river were left fallow during the flood season in readiness to receive the waters of the flood and to give rise to a larger tract of arable land.

The disastrous floods dreaded in those



**Figure 2** A pitcher of *Nepenthes albomarginata* being visited by termites. Nasuti termite workers and soldiers are harvesting the white rim hairs, which may mimic the appearance of termites' food. As the insects lose their grip on the peristome, hundreds or even thousands fall into the pitcher, where they are trapped.

days were those that lasted for long after they were supposed to recede, and which reached heights of a metre or more above that of AD 741/742. Only at that height would the flood reach the ground upon which most, if not all, the towns and villages of Egypt were built. Such dangerous floods were common during the fourteenth century and have been singled out by the historians, who dwelt at length on their devastating effects. The readings of the Roda nilometer have been widely scrutinized by scientists and I know of no one who classified the flood of AD 741/742 as exceptionally high or devastating. This conclusion is confirmed by a review of the earlier<sup>5,6</sup> and more recent literature<sup>7,8</sup>. I believe, therefore, that Stanley *et al.* need to re-examine their data in the light of this evidence.

#### Rushdi Said

Former Head of The Geological Survey of Egypt,  
3801 Millcreek Drive, Annandale,  
Virginia 22003, USA  
e-mail: rushsaid@hotmail.com

1. Stanley, J.-D. *et al.* *Nature* **412**, 293–294 (2001).
2. Nur, A. M. *EOS, Trans. Am. Geophys. Union* **81**, F21 (2000).
3. Pain, S. *New Scientist* 43–45 (20 October 2001).
4. Toussoun, O. *Mem. Inst. Egypte* **4**, 1–212 (1922).
5. Popper, W. *The Cairo Nilometer* (Univ. California Press, Berkeley, 1951).
6. Said, R. *The River Nile, Geology, Hydrology and Utilization* (Pergamon, Oxford, 1993).
7. Reussing, G. *Berliner Geowissenschaftliche Abh. D* **7** (1994).
8. Fraedrich, K. *et al.* *Int. J. Climatol.* **17**, 1301–1315 (1997).

#### Palaeobotany

## Atmospheric CO<sub>2</sub> from fossil plant cuticles

Plants respond to changes in atmospheric carbon dioxide levels by regulating the number of stomata in their leaves. In his reconstruction of a continuous, 300-million-year record of atmospheric CO<sub>2</sub>, Retallack bases his curve on stomatal counts of fossil plant cuticles taken from published micrographs<sup>1</sup>. However, the preservation of cuticles from Permian times is generally too fragmentary for the stomatal index to be reliably determined, the micrographs used could have biased the results, and there are important errors in the supplementary data<sup>1</sup> — all of which cast doubt on the Permian part of Retallack's record.

Not only do the fragmentary preservation of Permian plant cuticles and the small number of specimens counted call into question the statistical validity of Retallack's stomatal index, but the record may also be biased by reliance on cuticle micrographs that are not representative. In many species, stomata are not evenly distributed — for example, *Autunia conferta* and *Peltaspermum retensorium* have very different abaxial (lower) and adaxial (upper) leaf cuticles, particularly with regard to stomata distribution<sup>2–5</sup>. The

thinner abaxial cuticles have many more stomata, but their cell patterns can be indistinct (the positions of stomata are indicated by papillae on the subsidiary cells).

Most of the micrographs used by Retallack show adaxial cuticles with stomata concentrated between the veins, notably in the central part of the pinnules (leaflets); stomata are rare or absent on the rest of the adaxial side. In *P. retensorium*, stomata are present only in the basal part of the pinnules<sup>5</sup>; over 90% of the adaxial surface lacks stomata. The photographs used by Retallack mostly show stomata-bearing pieces of cuticle because stomata are of primary importance for taxonomy; stomata-poor or stomata-free cuticles are rarely shown in such images. Counting only these cuticles may lead to unreliable estimates of stomatal indices. Substantial variations in stomatal index have been reported in extant plants and such data should be interpreted with caution<sup>6,7</sup>.

Retallack's data<sup>1</sup> for the Permian contain several errors: stomatal counts are given for locations that have yielded no cuticle (for example, Sobernheim and Saxony '*Autunia conferta*'), and for material not illustrated in this context (Lebach: *A. conferta*). Ages given for some Permian localities are doubtful (Sobernheim is given as 280 ± 3 Myr (basal part of the Nahe Group (N 4); ref. 8), which is too young: the Grenzlagervolcanism that immediately underlies the Nahe Group has been dated (by Rb–Sr dating) at 290.7 ± 0.9 Myr; ref. 9); Frankenberg (253 ± 2 Myr) and Geismar (256 ± 2 Myr) are different designations<sup>1</sup> for the same locality, which is of the same age as the British Zechstein localities (Middridge, Kimberly, Cinderhill: 258 ± 2 Myr). Also, Lebach (Sakmarian: 285 ± 3 Myr) and Rümmlbach (Sakmarian–Artinskian: 283 ± 3 Myr) are outcrops in the same horizon (top Lauterecken–Oderheim Formation (L–O 10); ref. 8), underlying the Grenzlagervolcanics<sup>9</sup>, so Retallack's age for Lebach/Rümmlbach is too young). The actual record is thus much more punctuated than proposed<sup>1</sup>.

Expansion of the atmospheric-CO<sub>2</sub> curve into the Palaeozoic era is important, but this should be based on critical evaluation of more reliable data.

#### Hans Kerp

Forschungsstelle für Paläobotanik,  
Geologisch-Paläontologisches Institut, Westfälische  
Wilhelms-Universität, 48143 Münster, Germany  
e-mail: kerp@uni-muenster.de

1. Retallack, G. J. *Nature* **411**, 287–290 (2001).
2. Barthel, M. & Haubold, H. *Schriftenr. Geol. Wiss.* **16**, 49–105 (1980).
3. Kerp, J. H. F. *Rev. Palaeobot. Palynol.* **54**, 249–360 (1988).
4. Kerp, H. & Barthel, M. *Rev. Palaeobot. Palynol.* **78**, 1–18 (1993).
5. Naugolnykh, S. V. & Kerp, H. *Rev. Palaeobot. Palynol.* **91**, 35–62 (1996).
6. Poole, I. *et al.* *Plant Cell Environ.* **19**, 705–712 (1996).
7. Poole, I. *et al.* *New Phytol.* **145**, 511–521 (2000).
8. Boy, J. A. & Fichter, J. Z. *Dt. Geol. Ges.* **133**, 607–642 (1982).
9. Lippolt, H. J. & Hess, J. C. N. *Jb. Geol. Paläont. Mh.* 1989, 553–559 (1989).

*Retallack replies* — My data on the stomatal density of fossil plants through time<sup>1</sup> was made available to encourage refinements such as that now offered by Kerp. His corrections make no significant difference to my published curves, but future refinement should improve this palaeobotanical archive of atmospheric CO<sub>2</sub> levels.

Kerp's characterization of stomatal distribution on Permian seed-fern leaves is similar to the situation in *Lepidopteris stormbergensis*, which has a highly variable stomatal index<sup>2,3</sup>. My rarefaction analysis of several fossil species, including those described by Kerp, shows *L. stormbergensis* to be the most variable taxon of my compilation; I therefore used this species to set the lower boundary for reliable analyses at 500 epidermal cells<sup>1</sup>. The stomatal index of living *Ginkgo biloba* can be determined reliably by counting as few as 50 cells; other species represented by the thousands of cells needed for rarefaction analysis fell between these extremes.

Differences between Kerp's taxonomic names and mine reflect a different view of palaeobotanical nomenclature. He gives fossil leaves the same name as reproductive structures that are considered on evidence of varying quality to have belonged to the same plant. This is risky, because few, if any, of the taxa studied have reproductive structures attached to leaves. In my compilation of fossil leaves<sup>1</sup>, I listed names as he cited them for ease of reference, but in quotes to indicate deviation from traditional palaeobotanical form genera, such as the leaf genus *Rhachiphyllum*. Such taxonomic considerations do not affect the inferred CO<sub>2</sub> curve, because each determination is made using a collection of leaves at a single locality that are thought to belong to the same species of the same geological age.

I welcome Kerp's comments on local stratigraphic relationships. He has indicated that there are problems with fossil plant cuticles that he previously labelled as being from Sobernheim<sup>4</sup>, Lebach<sup>4</sup> and Saxony<sup>5</sup> (pointing out that the cuticle from Saxony is not '*Autunia conferta*', and '*Autunia conferta*' from 'Lebach' and 'Sobernheim' is really all from Langenthal, for which amended stomatal data are: stomatal index = 8.6 ± 1.1; Ne = 797, Ns = 80, Nf = 3). The best way forward is to count more cells from more fossils, at more localities tied to better-dated successions.

#### Gregory Retallack

Department of Geological Sciences, University of  
Oregon, Eugene, Oregon 97403, USA  
e-mail: gregr@darkwing.uoregon.edu

1. Retallack, G. J. *Nature* **411**, 287–290 (2001).
2. Townrow, J. A. *Palaeontology* **3**, 333–362 (1960).
3. Anderson, J. M. & Anderson, H. M. *Palaeoflora of Southern Africa, Molteno Formation (Triassic)*. Vol. 2 (Balkema, Rotterdam, 1989).
4. Kerp, J. H. F. *Rev. Palaeobot. Palynol.* **90**, 263–285 (1996).
5. Kerp, J. H. F. *Palaios* **5**, 548–569 (1990).



# Quantum phase transition from a superfluid to a Mott insulator in a gas of ultracold atoms

Markus Greiner\*, Olaf Mandel\*, Tilman Esslinger†, Theodor W. Hänsch\* & Immanuel Bloch\*

\* Sektion Physik, Ludwig-Maximilians-Universität, Schellingstrasse 4/III, D-80799 Munich, Germany, and Max-Planck-Institut für Quantenoptik, D-85748 Garching, Germany

† Quantenelektronik, ETH Zürich, 8093 Zurich, Switzerland

**For a system at a temperature of absolute zero, all thermal fluctuations are frozen out, while quantum fluctuations prevail. These microscopic quantum fluctuations can induce a macroscopic phase transition in the ground state of a many-body system when the relative strength of two competing energy terms is varied across a critical value. Here we observe such a quantum phase transition in a Bose–Einstein condensate with repulsive interactions, held in a three-dimensional optical lattice potential. As the potential depth of the lattice is increased, a transition is observed from a superfluid to a Mott insulator phase. In the superfluid phase, each atom is spread out over the entire lattice, with long-range phase coherence. But in the insulating phase, exact numbers of atoms are localized at individual lattice sites, with no phase coherence across the lattice; this phase is characterized by a gap in the excitation spectrum. We can induce reversible changes between the two ground states of the system.**

A physical system that crosses the boundary between two phases changes its properties in a fundamental way. It may, for example, melt or freeze. This macroscopic change is driven by microscopic fluctuations. When the temperature of the system approaches zero, all thermal fluctuations die out. This prohibits phase transitions in classical systems at zero temperature, as their opportunity to change has vanished. However, their quantum mechanical counterparts can show fundamentally different behaviour. In a quantum system, fluctuations are present even at zero temperature, due to Heisenberg's uncertainty relation. These quantum fluctuations may be strong enough to drive a transition from one phase to another, bringing about a macroscopic change.

A prominent example of such a quantum phase transition is the change from the superfluid phase to the Mott insulator phase in a system consisting of bosonic particles with repulsive interactions hopping through a lattice potential. This system was first studied theoretically in the context of superfluid-to-insulator transitions in liquid helium<sup>1</sup>. Recently, Jaksch *et al.*<sup>2</sup> have proposed that such a transition might be observable when an ultracold gas of atoms with repulsive interactions is trapped in a periodic potential. To illustrate this idea, we consider an atomic gas of bosons at low enough temperatures that a Bose–Einstein condensate is formed. The condensate is a superfluid, and is described by a wavefunction that exhibits long-range phase coherence<sup>3</sup>. An intriguing situation appears when the condensate is subjected to a lattice potential in which the bosons can move from one lattice site to the next only by tunnel coupling. If the lattice potential is turned on smoothly, the system remains in the superfluid phase as long as the atom–atom interactions are small compared to the tunnel coupling. In this regime a delocalized wavefunction minimizes the dominant kinetic energy, and therefore also minimizes the total energy of the many-body system. In the opposite limit, when the repulsive atom–atom interactions are large compared to the tunnel coupling, the total energy is minimized when each lattice site is filled with the same number of atoms. The reduction of fluctuations in the atom number on each site leads to increased fluctuations in the phase. Thus in the state with a fixed atom number per site phase coherence is lost. In addition, a gap in the excitation spectrum appears. The competition between two terms in the underlying hamiltonian

(here between kinetic and interaction energy) is fundamental to quantum phase transitions<sup>4</sup> and inherently different from normal phase transitions, which are usually driven by the competition between inner energy and entropy.

The physics of the above-described system is captured by the Bose–Hubbard model<sup>1</sup>, which describes an interacting boson gas in a lattice potential. The hamiltonian in second quantized form reads:

$$H = -J \sum_{\langle i,j \rangle} \hat{a}_i^\dagger \hat{a}_j + \sum_i \epsilon_i \hat{n}_i + \frac{1}{2} U \sum_i \hat{n}_i (\hat{n}_i - 1) \quad (1)$$

Here  $\hat{a}_i$  and  $\hat{a}_i^\dagger$  correspond to the bosonic annihilation and creation operators of atoms on the  $i$ th lattice site,  $\hat{n}_i = \hat{a}_i^\dagger \hat{a}_i$  is the atomic number operator counting the number of atoms on the  $i$ th lattice site, and  $\epsilon_i$  denotes the energy offset of the  $i$ th lattice site due to an external harmonic confinement of the atoms<sup>2</sup>. The strength of the tunnelling term in the hamiltonian is characterized by the hopping matrix element between adjacent sites  $i, j$ :  $J = -\int d^3x w(\mathbf{x} - \mathbf{x}_j) (-\hbar^2 \nabla^2 / 2m + V_{\text{lat}}(\mathbf{x})) w(\mathbf{x} - \mathbf{x}_i)$ , where  $w(\mathbf{x} - \mathbf{x}_i)$  is a single particle Wannier function localized to the  $i$ th lattice site (as long as  $n_i \approx O(1)$ ),  $V_{\text{lat}}(\mathbf{x})$  indicates the optical lattice potential and  $m$  is the mass of a single atom. The repulsion between two atoms on a single lattice site is quantified by the on-site interaction matrix element  $U = (4\pi\hbar^2 a/m) \int |w(\mathbf{x})|^4 d^3x$ , with  $a$  being the scattering length of an atom. In our case the interaction energy is very well described by the single parameter  $U$ , due to the short range of the interactions, which is much smaller than the lattice spacing.

In the limit where the tunnelling term dominates the hamiltonian, the ground-state energy is minimized if the single-particle wavefunctions of  $N$  atoms are spread out over the entire lattice with  $M$  lattice sites. The many-body ground state for a homogeneous system ( $\epsilon_i = \text{const.}$ ) is then given by:

$$|\Psi_{\text{SF}}\rangle_{U=0} \propto \left( \sum_{i=1}^M \hat{a}_i^\dagger \right)^N |0\rangle \quad (2)$$

Here all atoms occupy the identical extended Bloch state. An important feature of this state is that the probability distribution

for the local occupation  $n_i$  of atoms on a single lattice site is poissonian, that is, its variance is given by  $\text{Var}(n_i) = \langle n_i \rangle$ . Furthermore, this state is well described by a macroscopic wavefunction with long-range phase coherence throughout the lattice.

If interactions dominate the hamiltonian, the fluctuations in atom number of a Poisson distribution become energetically very costly and the ground state of the system will instead consist of localized atomic wavefunctions with a fixed number of atoms per site that minimize the interaction energy. The many-body ground state is then a product of local Fock states for each lattice site. In this limit, the ground state of the many-body system for a commensurate filling of  $n$  atoms per lattice site in the homogeneous case is given by:

$$|\Psi_{\text{MI}}\rangle_{J=0} \propto \prod_{i=1}^M (\hat{a}_i^\dagger)^n |0\rangle \quad (3)$$

This Mott insulator state cannot be described by a macroscopic wavefunction like in a Bose condensed phase, and thus is not amenable to a treatment via the Gross-Pitaevskii equation or Bogoliubov's theory of weakly interacting bosons. In this state no phase coherence is prevalent in the system, but perfect correlations in the atom number exist between lattice sites.

As the strength of the interaction term relative to the tunnelling term in the Bose–Hubbard hamiltonian is changed, the system reaches a quantum critical point in the ratio of  $U/J$ , for which the system will undergo a quantum phase transition from the superfluid state to the Mott insulator state. In three dimensions, the phase transition for an average number of one atom per lattice site is expected to occur at  $U/J = z \times 5.8$  (see refs 1, 5, 6, 7), with  $z$  being the number of next neighbours of a lattice site. The qualitative change in the ground-state configuration below and above the quantum critical point is also accompanied by a marked change in the excitation spectrum of the system. In the superfluid regime, the excitation spectrum is gapless whereas the Mott insulator phase exhibits a gap in the excitation spectrum<sup>5–8</sup>. An essential feature of a quantum phase transition is that this energy gap  $\Delta$  opens up as the quantum critical point is crossed.

Studies of the Bose–Hubbard hamiltonian have so far included granular superconductors<sup>9,10</sup> and one- and two-dimensional Josephson junction arrays<sup>11–16</sup>. In the context of ultracold atoms, atom number squeezing has very recently been demonstrated with a Bose–Einstein condensate in a one-dimensional optical lattice<sup>17</sup>. The above experiments were mainly carried out in the limit of large boson occupancies  $n_i$  per lattice site, for which the problem can be well described by a chain of Josephson junctions.

In our present experiment we load  $^{87}\text{Rb}$  atoms from a Bose–Einstein condensate into a three-dimensional optical lattice potential. This system is characterized by a low atom occupancy per lattice site of the order of  $\langle n_i \rangle \approx 1 - 3$ , and thus provides a unique testing ground for the Bose–Hubbard model. As we increase the lattice potential depth, the hopping matrix element  $J$  decreases exponentially but the on-site interaction matrix element  $U$  increases. We are thereby able to bring the system across the critical ratio in  $U/J$ , such that the transition to the Mott insulator state is induced.

## Experimental technique

The experimental set-up and procedure to create  $^{87}\text{Rb}$  Bose–Einstein condensates are similar to those in our previous experimental work<sup>18,19</sup>. In brief, spin-polarized samples of laser-cooled atoms in the ( $F = 2$ ,  $m_F = 2$ ) state are transferred into a cigar-shaped magnetic trapping potential with trapping frequencies of  $\nu_{\text{radial}} = 240$  Hz and  $\nu_{\text{axial}} = 24$  Hz. Here  $F$  denotes the total angular momentum and  $m_F$  the magnetic quantum number of the state. Forced radio-frequency evaporation is used to create Bose–Einstein condensates with up to  $2 \times 10^5$  atoms and no discernible thermal component. The radial trapping frequencies are then relaxed over a

period of 500 ms to  $\nu_{\text{rad}} = 24$  Hz such that a spherically symmetric Bose–Einstein condensate with a Thomas–Fermi diameter of  $26 \mu\text{m}$  is present in the magnetic trapping potential.

In order to form the three-dimensional lattice potential, three optical standing waves are aligned orthogonal to each other, with their crossing point positioned at the centre of the Bose–Einstein condensate. Each standing wave laser field is created by focusing a laser beam to a waist of  $125 \mu\text{m}$  at the position of the condensate. A second lens and a mirror are then used to reflect the laser beam back onto itself, creating the standing wave interference pattern. The lattice beams are derived from an injection seeded tapered amplifier and a laser diode operating at a wavelength of  $\lambda = 852$  nm. All beams are spatially filtered and guided to the experiment using optical fibres. Acousto-optical modulators are used to control the intensity of the lattice beams and introduce a frequency difference of about 30 MHz between different standing wave laser fields. The polarization of a standing wave laser field is chosen to be linear and orthogonal polarized to all other standing waves. Due to the different frequencies in each standing wave, any residual interference between beams propagating along orthogonal directions is time-averaged to zero and therefore not seen by the atoms. The resulting three-dimensional optical potential (see ref. 20 and references therein) for the atoms is then proportional to the sum of the intensities of the three standing waves, which leads to a simple cubic type geometry of the lattice:

$$V(x, y, z) = V_0(\sin^2(kx) + \sin^2(ky) + \sin^2(kz)) \quad (4)$$

Here  $k = 2\pi/\lambda$  denotes the wavevector of the laser light and  $V_0$  is the maximum potential depth of a single standing wave laser field. This depth  $V_0$  is conveniently measured in units of the recoil energy  $E_r = \hbar^2 k^2 / 2m$ . The confining potential for an atom on a single lattice site due to the optical lattice can be approximated by a harmonic potential with trapping frequencies  $\nu_r$  on the order of  $\nu_r \approx (\hbar k^2 / 2\pi m) \sqrt{V_0/E_r}$ . In our set-up potential depths of up to  $22 E_r$  can be reached, resulting in trapping frequencies of approximately  $\nu_r \approx 30$  kHz. The gaussian intensity profile of the laser beams at the position of the condensate creates an additional weak isotropic harmonic confinement over the lattice, with trapping frequencies of 65 Hz for a potential depth of  $22 E_r$ .

The magnetically trapped condensate is transferred into the optical lattice potential by slowly increasing the intensity of the lattice laser beams to their final value over a period of 80 ms using an exponential ramp with a time constant of  $\tau = 20$  ms. The slow ramp speed ensures that the condensate always remains in the many-body ground state of the combined magnetic and optical trapping potential. After raising the lattice potential the condensate has been distributed over more than 150,000 lattice sites ( $\sim 65$  lattice sites in a single direction) with an average atom number of up to 2.5 atoms per lattice site in the centre.

In order to test whether there is still phase coherence between different lattice sites after ramping up the lattice potential, we suddenly turn off the combined trapping potential. The atomic wavefunctions are then allowed to expand freely and interfere with each other. In the superfluid regime, where all atoms are delocalized over the entire lattice with equal relative phases between different lattice sites, we obtain a high-contrast three-dimensional interference pattern as expected for a periodic array of phase coherent matter wave sources (see Fig. 1). It is important to note that the sharp interference maxima directly reflect the high degree of phase coherence in the system for these experimental values.

## Entering the Mott insulator phase

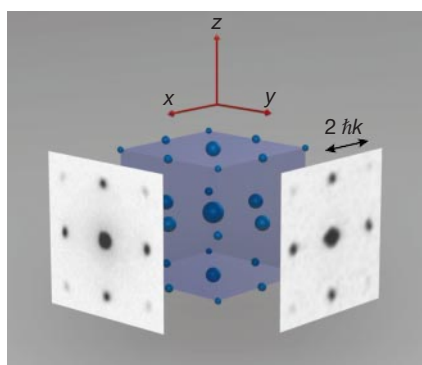
As we increase the lattice potential depth, the resulting interference pattern changes markedly (see Fig. 2). Initially the strength of higher-order interference maxima increases as we raise the potential height, due to the tighter localization of the atomic wavefunctions at a single lattice site. Quite unexpectedly, however, at a potential



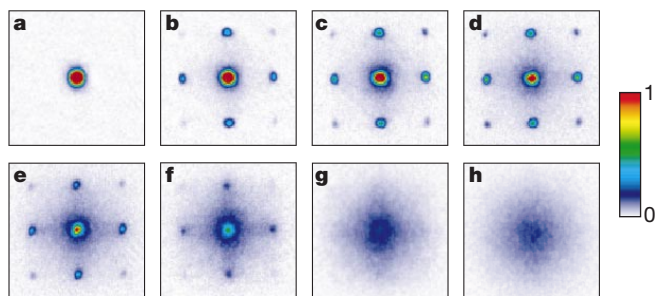
depth of around  $13 E_r$  the interference maxima no longer increase in strength (see Fig. 2e): instead, an incoherent background of atoms gains more and more strength until at a potential depth of  $22 E_r$  no interference pattern is visible at all. Phase coherence has obviously been completely lost at this lattice potential depth. A remarkable feature during the evolution from the coherent to the incoherent state is that when the interference pattern is still visible no broadening of the interference peaks can be detected until they completely vanish in the incoherent background. This behaviour can be explained on the basis of the superfluid–Mott insulator phase diagram. After the system has crossed the quantum critical point  $U/J = z \times 5.8$ , it will evolve in the inhomogeneous case into alternating regions of incoherent Mott insulator phases and coherent superfluid phases<sup>2</sup>, where the superfluid fraction continuously decreases for increasing ratios  $U/J$ .

### Restoring coherence

A notable property of the Mott insulator state is that phase coherence can be restored very rapidly when the optical potential is lowered again to a value where the ground state of the many-body system is completely superfluid. This is shown in Fig. 3. After only 4 ms of ramp-down time, the interference pattern is fully visible again, and after 14 ms of ramp-down time the interference peaks have narrowed to their steady-state value, proving that phase coherence has been restored over the entire lattice. The timescale for the restoration of coherence is comparable to the tunnelling time  $\tau_{\text{tunnel}} = \hbar/J$  between two neighbouring lattice sites in the system,



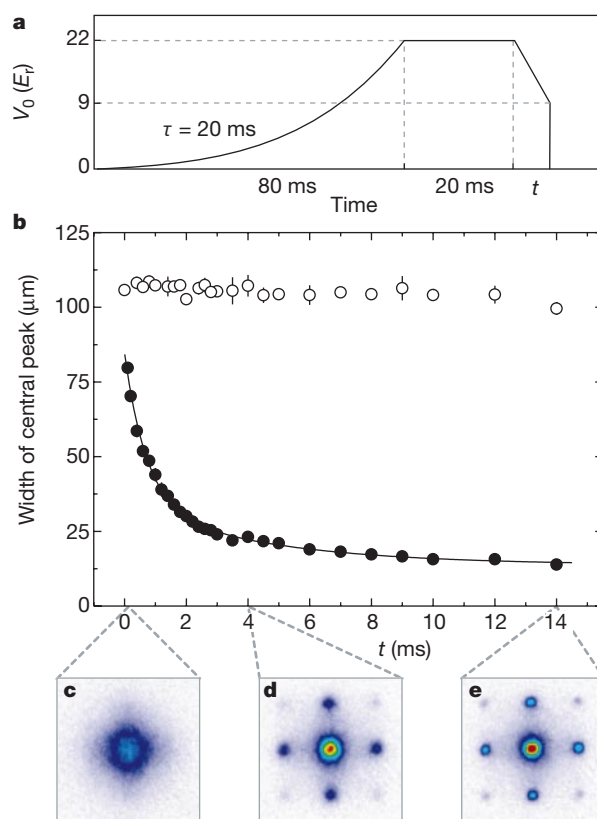
**Figure 1** Schematic three-dimensional interference pattern with measured absorption images taken along two orthogonal directions. The absorption images were obtained after ballistic expansion from a lattice with a potential depth of  $V_0 = 10 E_r$  and a time of flight of 15 ms.



**Figure 2** Absorption images of multiple matter wave interference patterns. These were obtained after suddenly releasing the atoms from an optical lattice potential with different potential depths  $V_0$  after a time of flight of 15 ms. Values of  $V_0$  were: **a**,  $0 E_r$ ; **b**,  $3 E_r$ ; **c**,  $7 E_r$ ; **d**,  $10 E_r$ ; **e**,  $13 E_r$ ; **f**,  $14 E_r$ ; **g**,  $16 E_r$ ; and **h**,  $20 E_r$ .

which is of the order of 2 ms for a lattice with a potential depth of  $9 E_r$ . A significant degree of phase coherence is thus already restored on the timescale of a tunnelling time.

It is interesting to compare the rapid restoration of coherence coming from a Mott insulator state to that of a phase incoherent state, where random phases are present between neighbouring lattice sites and for which the interference pattern also vanishes. This is shown in Fig. 3b, where such a phase incoherent state is created during the ramp-up time of the lattice potential (see Fig. 3 legend) and where an otherwise identical experimental sequence is used. Such phase incoherent states can be clearly identified by adiabatically mapping the population of the energy bands onto the Brillouin zones<sup>19,21</sup>. When we turn off the lattice potential adiabatically, we find that a statistical mixture of states has been created, which homogeneously populates the first Brillouin zone of

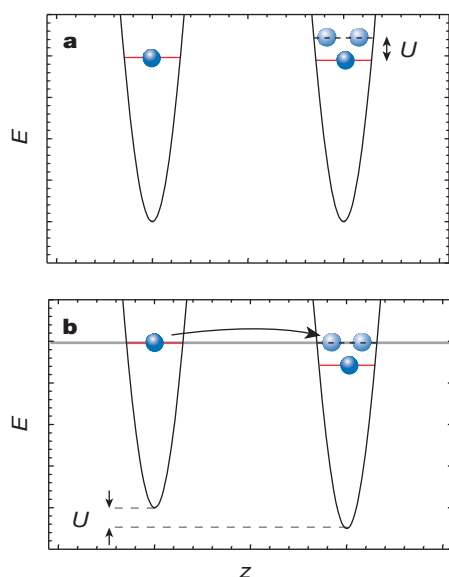


**Figure 3** Restoring coherence. **a**, Experimental sequence used to measure the restoration of coherence after bringing the system into the Mott insulator phase at  $V_0 = 22 E_r$  and lowering the potential afterwards to  $V_0 = 9 E_r$ , where the system is superfluid again. The atoms are first held at the maximum potential depth  $V_0$  for 20 ms, and then the lattice potential is decreased to a potential depth of  $9 E_r$  in a time  $t$  after which the interference pattern of the atoms is measured by suddenly releasing them from the trapping potential. **b**, Width of the central interference peak for different ramp-down times  $t$ , based on a lorentzian fit. In case of a Mott insulator state (filled circles) coherence is rapidly restored already after 4 ms. The solid line is a fit using a double exponential decay ( $\tau_1 = 0.94(7)$  ms,  $\tau_2 = 10(5)$  ms). For a phase incoherent state (open circles) using the same experimental sequence, no interference pattern reappears again, even for ramp-down times  $t$  of up to 400 ms. We find that phase incoherent states are formed by applying a magnetic field gradient over a time of 10 ms during the ramp-up period, when the system is still superfluid. This leads to a dephasing of the condensate wavefunction due to the nonlinear interactions in the system. **c–e**, Absorption images of the interference patterns coming from a Mott insulator phase after ramp-down times  $t$  of 0.1 ms (**c**), 4 ms (**d**), and 14 ms (**e**).

the three-dimensional lattice. This homogeneous population proves that all atoms are in the vibrational ground state of the lattice, but the relative phase between lattice sites is random. Figure 3b shows that no phase coherence is restored at all for such a system over a period of 14 ms. Even for evolution times  $t$  of up to 400 ms, no reappearance of an interference pattern could be detected. This demonstrates that the observed loss of coherence with increasing potential depth is not simply due to a dephasing of the condensate wavefunction.

### Probing the excitation spectrum

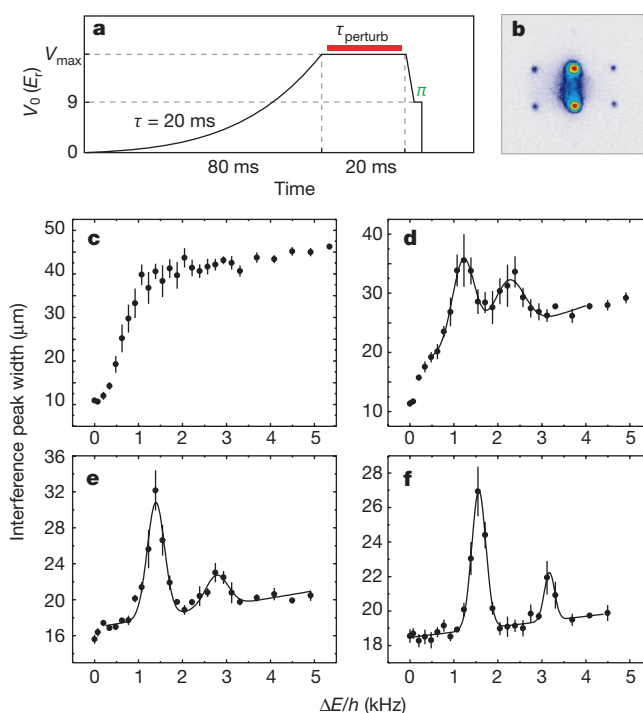
In the Mott insulator state, the excitation spectrum is substantially modified compared to that of the superfluid state. The excitation spectrum has now acquired an energy gap  $\Delta$ , which in the limit  $J \ll U$  is equal to the on-site interaction matrix element  $\Delta = U$  (see refs 5–8). This can be understood within a simplified picture in the following way. We consider a Mott insulator state with exactly  $n = 1$  atom per lattice site. The lowest lying excitation for such a state is the creation of a particle–hole pair, where an atom is removed from a lattice site and added to a neighbouring lattice site (see Fig. 4a). Due to the on-site repulsion between two atoms, the energy of the state describing two atoms in a single lattice site is raised by an amount  $U$  in energy above the state with only a single atom in this lattice site. Therefore in order to create an excitation the finite amount of energy  $U$  is required. It can be shown that this is also true for number states with exactly  $n$  atoms per lattice site. Here the energy required to make a particle–hole excitation is also  $U$ . Hopping of particles throughout the lattice is therefore suppressed in the Mott insulator phase, as this energy is only available in virtual processes. If now the lattice potential is tilted by application of a potential gradient, tunnelling is allowed again if the energy difference between neighbouring lattice sites due to the potential gradient equals the on-site interaction energy  $U$  (see Fig. 4b). We thus expect a resonant excitation probability versus the applied



**Figure 4** Excitation gap in the Mott insulator phase with exactly  $n = 1$  atom per lattice site. **a**, The lowest lying excitations in the Mott insulator phase consist of removing an atom from a lattice site and adding it to neighbouring lattice sites. Owing to the on-site repulsion between the atoms, this requires a finite amount  $U$  in energy and hopping of the atoms is therefore suppressed. **b**, If a potential gradient is applied to the system along the  $z$ -direction, such that the energy difference between neighbouring lattice sites equals the on-site interaction energy  $U$ , atoms are allowed to tunnel again. Particle–hole excitations are then created in the Mott insulator phase.

energy difference between neighbouring lattice sites for a Mott insulator phase.

We probe this excitation probability by using the experimental sequence shown in Fig. 5a. If excitations have been created during the application of the potential gradient at the potential depth  $V_0 = V_{\max}$ , we will not be able to return to a perfectly coherent superfluid state by subsequently lowering the potential to a depth of  $V_0 = 9E_r$ . Instead, excitations in the Mott insulator phase will lead to excitations in the lowest energy band in the superfluid case. These excitations are simply phase fluctuations between lattice sites, and cause a broadening of the interference maxima in the interference pattern (see Fig. 5b). Figure 5c–f shows the width of the interference peaks versus the applied gradient for four different potential depths  $V_{\max}$ . For a completely superfluid system at  $10 E_r$ , the system is easily perturbed already for small potential gradients and for stronger gradients a complete dephasing of the wavefunctions leads to a saturation in the width of the interference peaks. At a potential depth of about  $13 E_r$  two broad resonances start to appear in the excitation spectrum, and for a potential depth of  $20 E_r$  a dramatic



**Figure 5** Probing the excitation probability versus an applied vertical potential gradient. **a**, Experimental sequence. The optical lattice potential is increased in 80 ms to a potential depth  $V_0 = V_{\max}$ . Then the atoms are held for a time of 20 ms at this potential depth, during which a potential gradient is applied for a time  $\tau_{\text{perturb}}$ . The optical potential is then lowered again within 3 ms to a value of  $V_0 = 9E_r$ , for which the system is superfluid again. Finally, a potential gradient is applied for 300  $\mu\text{s}$  with a fixed strength, such that the phases between neighbouring lattice sites in the vertical direction differ by  $\pi$ . The confining potential is then rapidly turned off and the resulting interference pattern is imaged after a time of flight of 15 ms (**b**). Excitations created by the potential gradient at a lattice depth of  $V_0 = V_{\max}$  will lead to excitations in the superfluid state at  $V_0 = 9E_r$ . Here excitations correspond to phase fluctuations across the lattice, which will influence the width of the observed interference peaks. **c–f**, Width of interference peaks versus the energy difference between neighbouring lattice sites  $\Delta E$ , due to the potential gradient applied for a time  $\tau_{\text{perturb}}$ . **c**,  $V_{\max} = 10E_r$ ,  $\tau_{\text{perturb}} = 2$  ms; **d**,  $V_{\max} = 13E_r$ ,  $\tau_{\text{perturb}} = 6$  ms; **e**,  $V_{\max} = 16E_r$ ,  $\tau_{\text{perturb}} = 10$  ms; and **f**,  $V_{\max} = 20E_r$ ,  $\tau_{\text{perturb}} = 20$  ms. The perturbation times  $\tau_{\text{perturb}}$  have been prolonged for deeper lattice potentials in order to account for the increasing tunnelling times. The solid lines are fits to the data based on two gaussians on top of a linear background.



change in the excitation spectrum has taken place. Two narrow resonances are now clearly visible on top of an otherwise completely flat excitation probability. The slightly higher offset of the excitation probability for a deep optical lattice (Fig. 5e, f) compared to the initial width of the interference peaks in Fig. 5c, is due to the fact that after 3 ms ramp down time from a deep optical lattice, the system is still in the dynamical process of restoring coherence coming from the Mott insulator phase. For longer hold times this offset approaches almost the same initial width as in Fig. 5c, showing that we are not able to excite the system at all except for the two resonance gradients. At these large potential depths, the narrow resonances show that the energy gap  $\Delta$  of the system, which is measured here as the minimum energy difference between neighbouring lattice sites for which the system can be perturbed, is almost equal to the centre position of the resonance.

We have in fact found the Mott insulator state to be extremely robust to external perturbations, such as a modulation of the trapping potential or a modulation of the gradient potential, as long as the resonance gradients are avoided. The first resonance can be directly attributed to the creation of single particle-hole excitations in the Mott insulator state, and directly proves that we have indeed entered the Mott insulator regime. The second, weaker resonance occurs at exactly twice the energy difference of the first, stronger resonance. It can most probably be attributed to at least one of the following processes: (1) simultaneous tunnelling of two particles in a Mott insulator phase with  $n > 1$  atoms, (2) second-order processes, in which two particle-hole pairs are created simultaneously, with only one in the direction of the applied gradient, and (3) tunnelling processes occurring between lattice sites with  $n = 1$  atom next to lattice sites with  $n = 2$  atoms. In comparison, a two-dimensional lattice at a maximum potential depth of  $V_{\text{max}} = 20E_r$ , which we still expect to be in the superfluid regime, shows no resonances but a smooth excitation spectrum, similar to Fig. 5c.

The position of the resonances in the three-dimensional lattice can be seen to shift with increasing potential depth due to the tighter localization of the wave packets on a lattice site (see Fig. 6). We have compared the position of the first resonance versus the potential

depth  $V_{\text{max}}$  to an *ab initio* calculation of  $U$  based on Wannier functions from a band structure calculation, and find good agreement within our experimental uncertainties (see Fig. 6).

## Transition point

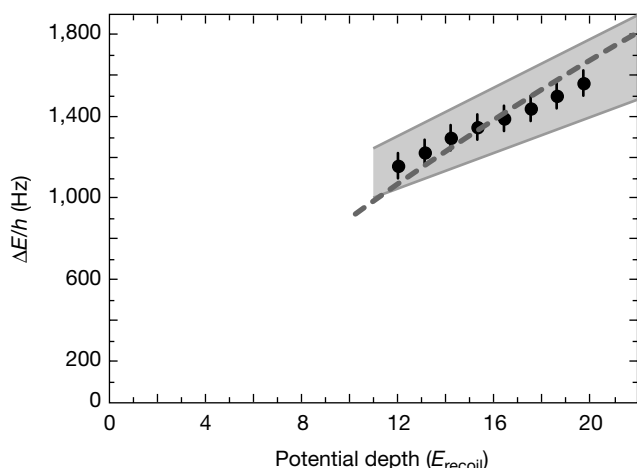
Both the vanishing of the interference pattern and the appearance of resonances in the excitation spectrum begin to occur at potential depths of  $V_0 = 12(1)–13(1) E_r$ , indicating the transition to the Mott insulator phase. We therefore expect the experimental transition point to lie above  $V_0 = 10(1) E_r$ , where no resonances are visible, and below  $V_0 = 13(1) E_r$ . It is important to compare this parameter range to the theoretical prediction based on the expected critical value  $U/J = z \times 5.8$ . In our simple cubic lattice structure, six next neighbours surround a lattice site.  $J$  and  $U$  can be calculated numerically from a band structure calculation for our experimental parameters, from which we find that  $U/J \approx 36$  for a potential depth of  $13 E_r$ . The theoretical prediction for the transition point is therefore in good agreement with the experimental parameter range for the transition point.

## Outlook

We have realized experimentally the quantum phase transition from a superfluid to a Mott insulator phase in an atomic gas trapped in an optical lattice. The experiment enters a new regime in the many-body physics of an atomic gas. This regime is dominated by atom-atom interactions and it is not accessible to theoretical treatments of weakly interacting gases, which have so far proved to be very successful in describing the physics of Bose-Einstein condensates<sup>22</sup>. The experimental realization of the Bose-Hubbard model with an atomic gas now allows the study of strongly correlated many-body quantum mechanics with unprecedented control of parameters. For example, besides controlling mainly the tunnelling matrix element, as done in this work, it should be possible in future experiments to control the atom-atom interactions via Feshbach resonances<sup>23,24</sup>.

The atoms in the Mott insulator phase can be considered as a new state of matter in atomic gases with unique properties. Atom number fluctuations at each lattice site are suppressed, and a well-defined phase between different lattices sites no longer exists. These number states have been proposed for the realization of a Heisenberg-limited atom interferometer<sup>25</sup>, which should be capable of achieving improved levels of precision. The Mott insulator phase also opens a new experimental avenue for recently proposed quantum gates with neutral atoms<sup>26</sup>. □

Received 26 October; accepted 29 November 2001.



**Figure 6** Energy difference between neighbouring lattice sites  $\Delta E$  for which the Mott insulator phase can be resonantly perturbed versus the lattice potential depth  $V_{\text{max}}$ . Experimental data points are shown as filled circles, and the shaded grey area denotes the possible variation of experimental values due to systematic uncertainties in the calibration of the potential depth and the applied gradient. The dashed line is the theoretical prediction for the on-site interaction matrix element  $U$ , based on Wannier functions from a band structure calculation.

1. Fisher, M. P. A., Weichman, P. B., Grinstein, G. & Fisher, D. S. Boson localization and the superfluid-insulator transition. *Phys. Rev. B* **40**, 546–570 (1989).
2. Jaksch, D., Bruder, C., Cirac, J. I., Gardiner, C. W. & Zoller, P. Cold bosonic atoms in optical lattices. *Phys. Rev. Lett.* **81**, 3108–3111 (1998).
3. Stringari, S. Bose-Einstein condensation and superfluidity in trapped atomic gases. *C.R. Acad. Sci.* **4**, 381–397 (2001).
4. Sachdev, S. *Quantum Phase Transitions* (Cambridge Univ. Press, Cambridge, 2001).
5. Sheshadri, K., Krishnamurthy, H. R., Pandit, R. & Ramakrishnan, T. V. Superfluid and insulating phases in an interacting-boson model: Mean-field theory and the RPA. *Europhys. Lett.* **22**, 257–263 (1993).
6. Freericks, J. K. & Monien, H. Phase diagram of the Bose Hubbard model. *Europhys. Lett.* **26**, 545–550 (1995).
7. van Oosten, D., van der Straten, P. & Stoof, H. T. C. Quantum phases in an optical lattice. *Phys. Rev. A* **63**, 053601-1–053601-12 (2001).
8. Elstner, N. & Monien, H. Dynamics and thermodynamics of the Bose-Hubbard model. *Phys. Rev. B* **59**, 12184–12187 (1999).
9. Orr, B. G., Jaeger, H. M., Goldman, A. M. & Kuper, C. G. Global phase coherence in two-dimensional granular superconductors. *Phys. Rev. Lett.* **56**, 378–381 (1986).
10. Haviland, D. B., Liu, Y. & Goldman, A. M. Onset of superconductivity in the two-dimensional limit. *Phys. Rev. Lett.* **62**, 2180–2183 (1989).
11. Bradley, R. M. & Doniach, S. Quantum fluctuations in chains of Josephson junctions. *Phys. Rev. B* **30**, 1138–1147 (1984).
12. Geerligs, L. J., Peters, M., de Groot, L. E. M., Verbruggen, A. & Mooij, J. E. Charging effects and quantum coherence in regular Josephson junction arrays. *Phys. Rev. Lett.* **63**, 326–329 (1989).
13. Zwerger, W. Global and local phase coherence in dissipative Josephson-junction arrays. *Europhys. Lett.* **9**, 421–426 (1989).

14. van der Zant, H. S. J., Fritschy, F. C., Elion, W. J., Geerligs, L. J. & Mooij, J. E. Field-induced superconductor-to-insulator transitions in Josephson-junction arrays. *Phys. Rev. Lett.* **69**, 2971–2974 (1992).
15. van Oudenaarden, A. & Mooij, J. E. One-dimensional Mott insulator formed by quantum vortices in Josephson junction arrays. *Phys. Rev. Lett.* **76**, 4947–4950 (1996).
16. Chow, E., Delsing, P. & Haviland, D. B. Length-scale dependence of the superconductor-to-insulator quantum phase transition in one dimension. *Phys. Rev. Lett.* **81**, 204–207 (1998).
17. Orzel, C., Tuchman, A. K., Fenselau, M. L., Yasuda, M. & Kasevich, M. A. Squeezed states in a Bose-Einstein condensate. *Science* **291**, 2386–2389 (2001).
18. Greiner, M., Bloch, I., Hänsch, T. W. & Esslinger, T. Magnetic transport of trapped cold atoms over a large distance. *Phys. Rev. A* **63**, 031401-1–031401-4 (2001).
19. Greiner, M., Bloch, I., Mandel, O., Hänsch, T. W. & Esslinger, T. Exploring phase coherence in a 2D lattice of Bose-Einstein condensates. *Phys. Rev. Lett.* **87**, 160405-1–160405-4 (2001).
20. Grimm, R., Weidemüller, M. & Ovchinnikov, Yu. B. Optical dipole traps for neutral atoms. *Adv. At. Mol. Opt. Phys.* **42**, 95–170 (2000).
21. Kastberg, A., Phillips, W. D., Rolston, S. L., Spreeuw, R. J. C. & Jessen, P. S. Adiabatic cooling of cesium to 700 nK in an optical lattice. *Phys. Rev. Lett.* **74**, 1542–1545 (1995).
22. Dalfovo, F. D., Giorgini, S., Pitaevskii, L. P. & Stringari, S. Theory of Bose-Einstein condensation in trapped gases. *Rev. Mod. Phys.* **71**, 463–512 (1999).
23. Inouye, S. *et al.* Observation of Feshbach resonances in a Bose-Einstein condensate. *Nature* **392**, 151–154 (1998).
24. Donley, E. A. *et al.* Dynamics of collapsing and exploding Bose-Einstein condensates. *Nature* **412**, 295–299 (2001).
25. Bouyer, P. & Kasevich, M. Heisenberg-limited spectroscopy with degenerate Bose-Einstein gases. *Phys. Rev. A* **56**, R1083–R1086 (1997).
26. Jaksch, D., Briegel, H.-J., Cirac, J. I., Gardiner, C. W. & Zoller, P. Entanglement of atoms via cold controlled collisions. *Phys. Rev. Lett.* **82**, 1975–1978 (1999).

## Acknowledgements

We thank W. Zwerger, H. Monien, I. Cirac, K. Burnett and Yu. Kagan for discussions. This work was supported by the DFG, and by the EU under the QUEST programme.

## Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to I.B. (e-mail: imb@mpq.mpg.de).

# p53 mutant mice that display early ageing-associated phenotypes

Stuart D. Tyner<sup>\*†</sup>, Sundaresan Venkatachalam<sup>†‡</sup>, Jene Choi<sup>‡</sup>, Stephen Jones<sup>§</sup>, Nader Ghebranious<sup>||</sup>, Herbert Igelmann<sup>¶</sup>, Xiongbin Lu<sup>‡</sup>, Gabrielle Soron<sup>‡</sup>, Benjamin Cooper<sup>‡</sup>, Cory Brayton<sup>#</sup>, Sang Hee Park<sup>☆</sup>, Timothy Thompson<sup>☆</sup>, Gerard Karsenty<sup>††</sup>, Allan Bradley<sup>††‡</sup> & Lawrence A. Donehower<sup>‡§§</sup>

<sup>\*</sup> Cell and Molecular Biology Program; <sup>‡</sup> Department of Molecular Virology and Microbiology; <sup>#</sup> Center for Comparative Medicine; <sup>☆</sup> Scott Department of Urology; <sup>††</sup> Department of Human and Molecular Genetics; and <sup>§§</sup> Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, Texas 77030, USA

<sup>§</sup> Department of Cell Biology, University of Massachusetts Medical Center, Worcester, Massachusetts 01655, USA

<sup>||</sup> Marshfield Medical Research Foundation, 1000 North Oak Avenue, Marshfield, Wisconsin 54449, USA

<sup>¶</sup> Heinrich-Pette-Institut für Experimentelle Virologie und Immunologie an der Universität Hamburg, Martinistrasse 52, D-20251 Hamburg, Germany

<sup>‡‡</sup> The Sanger Centre, Wellcome Trust Genome Campus, Cambridgeshire CB10 1SA, UK

<sup>†</sup> These authors contributed equally to this work

**The p53 tumour suppressor is activated by numerous stressors to induce apoptosis, cell cycle arrest, or senescence. To study the biological effects of altered p53 function, we generated mice with a deletion mutation in the first six exons of the p53 gene that express a truncated RNA capable of encoding a carboxy-terminal p53 fragment. This mutation confers phenotypes consistent with activated p53 rather than inactivated p53. Mutant ( $p53^{+/m}$ ) mice exhibit enhanced resistance to spontaneous tumours compared with wild-type ( $p53^{+/+}$ ) littermates. As  $p53^{+/m}$  mice age, they display an early onset of phenotypes associated with ageing. These include reduced longevity, osteoporosis, generalized organ atrophy and a diminished stress tolerance. A second line of transgenic mice containing a temperature-sensitive mutant allele of p53 also exhibits early ageing phenotypes. These data suggest that p53 has a role in regulating organismal ageing.**

The p53 tumour suppressor protein responds to a variety of cellular stresses, including DNA damage, hypoxia and aberrantly activated oncogenes, and may induce cell cycle arrest or apoptosis<sup>1–3</sup>. Studies on cultured primary cells have linked p53 activation to cellular replicative senescence—a terminal cell cycle arrest state reached after a finite number of cell divisions<sup>4</sup>. When cultured human diploid fibroblasts senesce, the activation state of p53 increases<sup>5–7</sup>. Introduction of viral oncoproteins, mutant p53 forms, or anti-p53 antibodies (which inactivate wild-type p53) delay senescence of human primary cells<sup>8–10</sup>. Introduction of oncogenic Ras into primary murine fibroblasts is initially mitogenic, but then induces a premature senescence dependent on wild-type p53 (ref. 11). These data are consistent with hypotheses that cellular replicative senescence serves as a mechanism of tumour suppression<sup>4,12,13</sup>.

Recent reports about genetically engineered mice with altered longevity phenotypes implicate p53 as a mediator of organismal senescence<sup>14–18</sup>. Late-generation telomerase-deficient mice exhibit shortened telomeres, decreased longevity and early senescence-related phenotypes<sup>14</sup>. Cells from these mice showed high levels of activated p53 (ref. 15). Crossing of the telomerase-deficient mice with p53-deficient mice resulted in rescue of some senescence-related phenotypes<sup>15</sup>. Early senescence phenotypes were also attributed to a Ku80-deficient mouse that was defective for double-stranded DNA break repair<sup>16</sup>. Embryonic fibroblasts from these mice showed a premature senescence that was rescued by p53 deficiency<sup>17</sup>.  $p66^{shc}$  null mice displayed 30% enhanced longevity compared with wild-type mice<sup>18</sup>.  $p66^{shc}$  mediates a stress response to reactive oxygen species, and cells from the deficient mice exhibited a defective p53 apoptotic response<sup>18</sup>.

A direct test of the involvement of p53 in the longevity of organisms could involve modulation of p53 levels in the mouse. We, and others, have generated p53-deficient mice<sup>19–21</sup>, but were prevented from studying ageing by early tumours. Conversely, attempts to generate mice that overexpress wild-type p53 have been unsuccessful, probably owing to deleterious effects of increased p53 on the developing embryo (A. Bernstein and J. Butel, personal communication). We report the serendipitous

development of a genetically altered mouse that may express a truncated form of p53 that augments wild-type p53 activity. Mice with this defective p53 allele (which we term the 'm' allele) are highly resistant to tumours and display early onset of phenotypes associated with ageing. A second line of transgenic mice containing multiple copies of a temperature-sensitive mutant p53 allele<sup>22</sup> also displayed some of the phenotypes of early ageing observed in  $p53^{+/m}$  mice. The observation of such early ageing phenotypes in two distinct p53 mutant lines implicates p53 as a regulator of organismal ageing.

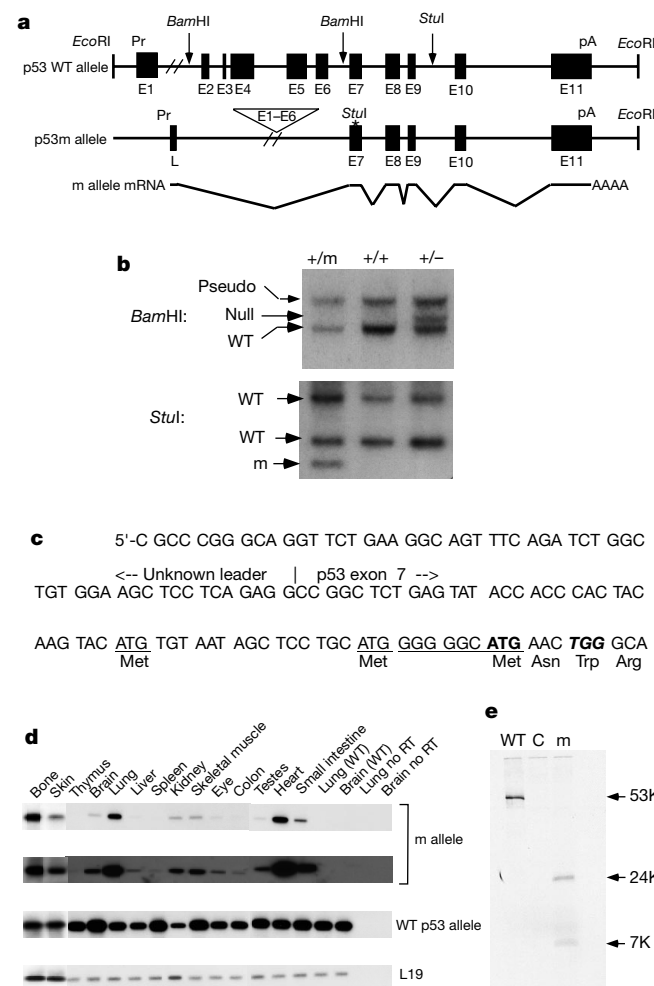
## Generation of $p53^{+/m}$ mice

The  $p53^{+/m}$  mouse was obtained following an aberrant gene-targeting event in embryonic stem cells. Initially, a p53 targeting construct containing a mutation at codon 245 (Arg245Trp) in p53 was introduced into embryonic stem cells (Fig. 1a). In one embryonic stem cell clone, the 3' segment of the targeting construct recombined with the endogenous p53 allele. Southern hybridization and polymerase chain reaction (PCR) assays revealed that the modified allele retained p53 exons 7–11 (including the codon 245 mutation) while deleting p53 exons 1–6 and an undefined region upstream of p53 (Fig. 1a, b). Preliminary mapping experiments indicate that the m-allele deletion extends over 20 kilobases (kb) upstream of p53. The deletion extends beyond the nearest p53 proximal gene, *Efnb3* (S.D.T. and X.L., unpublished data).

We generated mice containing the p53 m allele ( $p53^{+/m}$  mice). When the  $p53^{+/m}$  mice began to show unexpected phenotypes (see below), we determined that the p53 m allele was expressed as a messenger RNA. Using the 5' rapid amplification of complementary DNA ends (RACE) procedure, we identified a discrete amplified fragment in tissues of  $p53^{+/m}$  mice that was not found in tissues of  $p53^{+/+}$  mice. The sequence revealed a chimaeric message containing a 55-bp leader of unknown origin spliced into exon 7 of p53 (Fig. 1c). The foreign leader sequence contained no AUG codons, whereas the p53 exon 7 portion of the message contained three in-frame AUG codons, the third (Met codon 243) being adjacent to a viable translation initiation signal (Fig. 1c). Thus, the chimaeric



m-allele message could produce a C-terminal p53 fragment with a relative molecular mass of roughly 24,000 ( $M_r$  24K). To confirm this, transcripts of the m allele were generated by *in vitro* transcription and then translated by *in vitro* translation. As expected, a protein of 24K was produced that could be immunoprecipitated by a monoclonal antibody specific for the C terminus of p53 (Fig. 1e). However, to date we have been unable to detect the m protein in  $p53^{+/-}$  mouse tissues.



**Figure 1** Structure and expression of the mutant p53 m allele. **a**, Maps of the p53 wild-type (WT) and m alleles. Exons 1–6 and upstream sequences are deleted in the p53 m allele, juxtaposing exons 7–11 to an unknown promoter that initiates a chimaeric mRNA containing a 55-nucleotide leader (L) spliced into exons 7–11. Pr, promoter; pA, polyadenylation site. **b**, Southern analysis of tail DNAs from  $p53^{+/+}$ ,  $p53^{+/-}$  and  $p53^{-/-}$  mice. The top panel shows tail DNAs cleaved with BamHI and hybridized to a p53 exon 2–6 probe to differentiate p53 + and – alleles. The bottom panel shows DNAs cleaved with Stul and hybridized to a p53 exon 5–10 probe to distinguish p53 + and m alleles. **c**, Structure of m-allele mRNA. A 5' leader of unknown origin is spliced to the 5' boundary of p53 exon 7. This message contains three in-frame AUG codons. The third AUG codon is a viable translation initiation site (bold and underline), which may produce a C-terminal p53 fragment. The Arg245Trp codon change is indicated (italic and bold). **d**, RT-PCR analysis of tissues of  $p53^{+/+}$  and  $p53^{+/-}$  mice with primers specific for the m and wild-type p53 alleles. Top and second panel (15 $\times$  exposure of the top panel) show that most tissues of  $p53^{+/-}$  mice express the m-allele RNA, albeit at lower levels than wild-type p53 RNA (third panel). L19 RT-PCR reactions (fourth panel) provide controls. **e**, Protein-coding potential of p53 m-allele RNA. Constructs containing wild-type (WT) p53, no insert (C), or m-allele cDNA were transcribed and translated *in vitro*, and immunoprecipitated with a p53 C-terminal monoclonal antibody. The m-allele construct generated the predicted 24K protein.

To determine the expression pattern of p53 m-allele mRNA *in vivo*, we carried out PCR with reverse transcription (RT-PCR) assays specific for m-allele and wild-type p53 mRNAs. Most tissues of the  $p53^{+/-}$  mice, but not the  $p53^{+/+}$  mice, expressed m-allele mRNA, but at lower levels than that for wild-type p53 RNA (Fig. 1d).

## Enhanced tumour resistance and altered longevity

We monitored the  $p53^{+/-}$  mice for the presence of tumours over their lifespan. None of the 35  $p53^{+/-}$  mice developed overt, life-threatening tumours, whereas over 80% of  $p53^{+/+}$  mice and over 45% of  $p53^{+/-}$  mice developed such tumours (Table 1). On histopathological examination of the  $p53^{+/-}$  mice, 2 out of 35 were observed to have localized tumour lesions. One mouse had a focal lung adenoma and another had a well-differentiated bronchioalveolar lung adenocarcinoma (Table 1). In contrast,  $p53^{+/+}$  and  $p53^{-/-}$  mice developed large tumours of varied type, including lymphomas, osteosarcomas, soft tissue sarcomas and carcinomas.

Given the enhanced tumour resistance in  $p53^{+/-}$  mice, it might be expected that their longevity would exceed that of the more tumour-prone  $p53^{+/+}$  mice. In fact, the median lifespan of the  $p53^{+/-}$  mice was 96 weeks compared with 118 weeks for their  $p53^{+/+}$  littermates (Fig. 2a and Table 2). The maximal lifespan of the  $p53^{+/-}$  mice was 136 weeks compared with 164 weeks for the  $p53^{+/+}$  mice. Statistical comparison of the  $p53^{+/-}$  and  $p53^{+/+}$  survival curves indicated that the differences in longevity were highly significant ( $P < 0.0001$ ). Histopathological examination of the  $p53^{+/-}$  mice failed to reveal any obvious disease-associated phenotypes. A precise cause of death was difficult to determine in many cases. No inflammation or other indicators of infectious diseases above levels seen in  $p53^{+/+}$  mice were noted, although subtle differences in immune status or function cannot be ruled out.

To determine whether the altered longevity and tumour resistance in the  $p53^{+/-}$  mice were dependent on the presence of wild-type p53, we crossed  $p53^{+/-}$  mice to  $p53^{-/-}$  mice and obtained  $p53^{-/-}$  offspring. If the m-allele product augments the activity of wild-type p53 to enhance tumour resistance, then the absence of wild-type p53 in the  $p53^{-/-}$  mice should result in little or no tumour suppression effect when the survival curves of  $p53^{-/-}$  and  $p53^{+/+}$  mice are compared. Although there is a slight delay in mean tumorigenesis in  $p53^{-/-}$  mice compared with  $p53^{+/+}$  mice, maximal lifespans of the two genotypes are similar (9–10 months) (Fig. 2b). The tumour types observed in the  $p53^{-/-}$  mice were similar to those of the  $p53^{+/+}$  mice—primarily T-cell lymphomas and soft tissue sarcomas. Thus, the enhanced tumour resistance conferred by the m allele is dependent on the presence of wild-type p53.

## Augmented p53 responses in $p53^{+/-}$ mice

To assess mechanisms of tumour resistance in  $p53^{+/-}$  mice, we examined cells and tissues from  $p53^{+/+}$ ,  $p53^{+/-}$ ,  $p53^{-/-}$  and  $p53^{+/-}$  mice for p53-regulated activities. We compared  $p53^{+/+}$  and  $p53^{+/-}$  mouse embryonic fibroblasts (MEFs) for susceptibility to transformation after infection with an activated *ras* plus *myc* transforming retrovirus. We found that  $p53^{+/-}$  MEFs were consistently 2–4-fold more resistant to transformation than their  $p53^{+/+}$

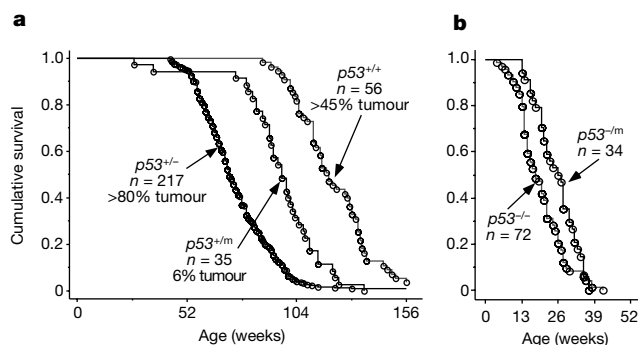
**Table 1** Tumour types arising in  $p53^{+/+}$ ,  $p53^{+/-}$ ,  $p53^{-/-}$  and pL53 mice

| Tumour type         | $p53^{+/+}$ (n = 217) | $p53^{+/-}$ (n = 56) | $p53^{-/-}$ (n = 35) | pL53 (n = 66) |
|---------------------|-----------------------|----------------------|----------------------|---------------|
| Lymphoma            | 49 (28%)              | 18 (67%)             | 0                    | 9 (56%)       |
| Osteosarcoma        | 61 (34%)              | 2 (7%)               | 0                    | 0             |
| Soft tissue sarcoma | 47 (27%)              | 3 (11%)              | 0                    | 1 (6%)        |
| Carcinoma           | 20 (11%)              | 4 (15%)              | 1 (50%)*             | 5 (31%)*      |
| Other               | 0                     | 0                    | 1 (50%)*             | 1 (6%)        |
| Total               | 177                   | 27                   | 2                    | 16            |

Values in parentheses indicate the percentage of total tumours.

\* Lung adenocarcinoma.

† Small focal lung adenoma.



**Figure 2** Longevity in  $p53^{+/+}$ ,  $p53^{+/-}$ ,  $p53^{+/m}$ ,  $p53^{-/-}$  and  $p53^{-/m}$  mice. **a**, Survival of  $p53^{+/-}$ ,  $p53^{+/+}$  and  $p53^{+/m}$  mice. More than half of the tumours in wild-type mice appeared while many of the  $p53^{+/m}$  mice were still alive. **b**, Survival of  $p53^{-/-}$  and  $p53^{-/m}$  mice. Virtually all of the  $p53^{-/-}$  and  $p53^{-/m}$  mice developed cancer.

counterparts (Fig. 3a). We also compared transformation efficiencies of  $p53^{-/-}$  and  $p53^{-/m}$  MEFs.  $p53^{-/m}$  cells displayed equivalent susceptibility to transformation as  $p53^{-/-}$  cells (Fig. 3a), indicating that transformation resistance mediated by the m allele is dependent on wild-type p53.

After DNA damage, levels of p53 protein increase and p53 activates a number of transcriptional targets, including the p21<sup>Waf1/Cip1</sup> cyclin-dependent kinase inhibitor. We compared p53 protein stability in kidneys of  $p53^{+/+}$ ,  $p53^{+/m}$  and  $p53^{+/-}$  mice treated with 5 Gy of ionizing radiation. Before irradiation, p53 protein levels were low, but p53 levels after normalization to an actin control showed a robust increase at 4 h after irradiation for kidney cells of  $p53^{+/+}$ ,  $p53^{+/-}$  and  $p53^{+/m}$  mice (Fig. 3b). However, by 24 h after irradiation, p53 protein levels in the cells from  $p53^{+/-}$  mice were decreased from those observed in  $p53^{+/+}$  and  $p53^{+/m}$  mice (Fig. 3b). These results indicate that the presence of the m allele may augment p53 protein stability.

p21<sup>Waf1/Cip1</sup> RNA was compared in MEFs of  $p53^{+/+}$ ,  $p53^{+/-}$  and  $p53^{+/m}$  mice before and after treatment with 5 Gy of ionizing radiation (Fig. 3c). Before irradiation, MEFs of  $p53^{+/+}$  and  $p53^{+/m}$  mice exhibited higher levels of basal p21 expression compared with MEFs of  $p53^{+/-}$  mice. At 4 h after ionizing radiation treatment, p21 mRNA levels were highest in MEFs of  $p53^{+/+}$  mice, with intermediate levels in MEFs of  $p53^{+/m}$  mice and relatively lower levels in MEFs of  $p53^{+/-}$  mice. By 48 h, levels of p21 remained high in cells of  $p53^{+/+}$  and  $p53^{+/m}$  mice, but declined in cells of  $p53^{+/-}$  mice.

The robust p21 expression in cells of  $p53^{+/m}$  mice was corroborated by transfection assays. Saos-2 human osteosarcoma cells (null

for p53) were co-transfected with a p21 promoter luciferase reporter construct and various combinations of m, wild-type p53, or empty vector expression constructs. Cells transfected with wild-type p53 or wild-type p53 plus empty vector constructs exhibited more than a tenfold increase in luciferase activity compared with cells transfected with empty vector or m expression constructs alone (Fig. 3d). In four independent experiments, wild-type p53 plus m-allele expression constructs induced a mean 2.3-fold increase in luciferase activity compared with cells transfected with wild-type p53 alone. Moreover, when the amount of the m expression construct was only one-tenth that of the co-transfected wild-type p53 construct, luciferase activity was still significantly higher (1.6-fold) than that produced by wild-type p53 alone. These data indicate that m-allele product can enhance p53 stability and transactivation activity.

### Early senescence-associated phenotypes in $p53^{+/m}$ mice

We examined the possibility that the shortened lifespans of  $p53^{+/m}$  mice could be accompanied by signs of early ageing. Up to 12 months of age  $p53^{+/m}$  mice appear morphologically identical to their  $p53^{+/+}$  littermates. However, by 18 months, virtually all of the  $p53^{+/m}$  mice exhibit signs of weight loss, lordokyphosis (hunchbacked spine) and an absence of vigour (Fig. 4a–d). Weighing the mice confirmed significant reductions in the body mass of  $p53^{+/m}$  mice at 18 and 24 months of age compared with their age-matched  $p53^{+/+}$  counterparts (Fig. 4d). In contrast, reduced body weights in the  $p53^{+/+}$  mice were not observed until extreme old age (30–36 months) (Fig. 4d). The weight losses of  $p53^{+/m}$  mice were not a result of reduced food intake. Young and old  $p53^{+/m}$  and  $p53^{+/+}$  animals consumed similar amounts of food. Moreover, histological examination of intestinal sections revealed no abnormalities consistent with a reduced ability to absorb nutrients. Blood chemistries of the older  $p53^{+/m}$  animals did not reveal any signs of nutritional deficiencies (S.T. and C.B., unpublished data; and data not shown).

Visual, pathological and histopathological examination of young (3 months)  $p53^{+/m}$  and  $p53^{+/+}$  mice failed to reveal any obvious differences between the two genotypes; however, assessment of old (24 months)  $p53^{+/m}$  and  $p53^{+/+}$  mice revealed marked differences. Skinned, older  $p53^{+/m}$  mice exhibit clear reductions in body mass, adipose tissue deposition, muscle mass and a pronounced lordokyphosis (Fig. 4c). The apparent differences in muscle mass were confirmed by weighing isolated quadriceps femoris muscles from 22–24-month-old  $p53^{+/m}$  and  $p53^{+/+}$  mice. The mean mass of the quadriceps of male  $p53^{+/+}$  mice was almost 2.5-fold that of the quadriceps of male  $p53^{+/m}$  mice (data not shown). Spleen, liver, kidney and testes, although not brain, were reduced in mass in  $p53^{+/m}$  mice (Fig. 4d–g). The reduced organ masses of  $p53^{+/m}$  mice are due to a decrease in overall cellularity. This is demonstrated in

**Table 2** Ageing-related phenotypes in  $p53^{+/+}$ ,  $p53^{+/m}$  and pL53 mice

| Phenotype                    | $p53^{+/+}$              | $p53^{+/m}$                 | pL53               |
|------------------------------|--------------------------|-----------------------------|--------------------|
| Median lifespan              | 118 weeks                | 96 weeks                    | ND                 |
| Maximum lifespan             | 164 weeks                | 136 weeks                   | ND                 |
| Cancer incidence             | >45%                     | <6%                         | 20% (18 months)    |
| Body weight                  | Reduced at 30 months     | Reduced at 18 months        | Slightly reduced   |
| Liver, spleen, kidney mass   | Minimal loss of mass     | 25–40% reduced at 24 months | Minimal loss       |
| Lymphoid atrophy             | Moderate                 | Pronounced                  | Moderate           |
| Lordokyphosis                | Modest                   | Pronounced                  | Pronounced         |
| Osteoporosis                 | Minimal                  | Pronounced                  | Pronounced         |
| Blood chemistry              | Normal                   | Normal                      | Normal             |
| Urinalysis                   | Normal                   | Normal                      | ND                 |
| Peripheral WBC, RBC          | Normal                   | Normal                      | Normal             |
| Hair greying and alopecia    | Minimal                  | Minimal                     | Some alopecia      |
| Hair regrowth                | Modestly reduced         | Greatly reduced             | Greatly reduced    |
| Dermal thickness             | Moderately reduced       | Greatly reduced             | Moderately reduced |
| Subcutaneous adipose         | Moderately reduced       | Greatly reduced             | Greatly reduced    |
| Wound-healing                | Normal                   | Retarded                    | Retarded           |
| Muscle atrophy               | Minimal                  | Pronounced                  | Minimal            |
| Anaesthetic stress tolerance | Well tolerated           | Poorly tolerated            | Poorly tolerated   |
| 5-FU myeloablation           | Robust WBC replenishment | Reduced WBC replenishment   | ND                 |

Phenotypes of  $p53^{+/+}$  and  $p53^{+/m}$  mice were assessed at 24 months of age; phenotypes of pL53 mice were assessed at 16–20 months of age. ND, not done; WBC, white blood cell; RBC, red blood cell.

the spleen sections, where the T- and B-cell-containing white pulp regions in 24-month-old  $p53^{+/m}$  mice are markedly reduced in area and cell number compared with those in the spleens of age-matched  $p53^{+/+}$  mice (Fig. 4h, i), indicative of a pronounced lymphoid atrophy. Both males and females showed comparable reductions in body weight and organ masses. These results correlate with human ageing, where after the age of 60, reductions in total body, liver, spleen and kidney masses are observed<sup>23</sup>.

The pronounced lordokyphosis in old  $p53^{+/m}$  mice suggested osteoporosis, an age-related marker in humans and mice<sup>24,25</sup>. X-ray analysis of 3-month-old mice failed to reveal any obvious differences in bone structure or density. By 12 months of age, some of the

$p53^{+/m}$  mice began to show global reductions in bone density, which became severe in the 24-month-old  $p53^{+/m}$  mice in comparison with age-matched  $p53^{+/+}$  littermates (Fig. 5a). The tail, vertebrae, limbs and craniums of the 24-month-old  $p53^{+/m}$  mice showed significant reductions in bone density. Histological analysis of cross-sections from 24-month-old  $p53^{+/m}$  and  $p53^{+/+}$  mice showed reduction in cortical and trabecular bone area compared with tibias of old  $p53^{+/+}$  mice (Fig. 5b, c).

Two markers for aged skin in humans are reduced dermal thickness and subcutaneous adipose<sup>26</sup>. Histological cross-sections of dorsal skin revealed no differences in the skin structure of 3-month-old mice, but significant differences were seen in the skin of 24-month-old mice (Fig. 6a–d). Skin of  $p53^{+/m}$  mice showed significant reductions in mean dermal thickness compared with skin of  $p53^{+/+}$  mice at 24 months (Fig. 6c–e). Although skin of young  $p53^{+/m}$  and  $p53^{+/+}$  mice had many subcutaneous adipose cells, these cells were virtually absent in skin of old  $p53^{+/m}$  mice and reduced in skin of old  $p53^{+/+}$  mice (Fig. 6c, d, f).

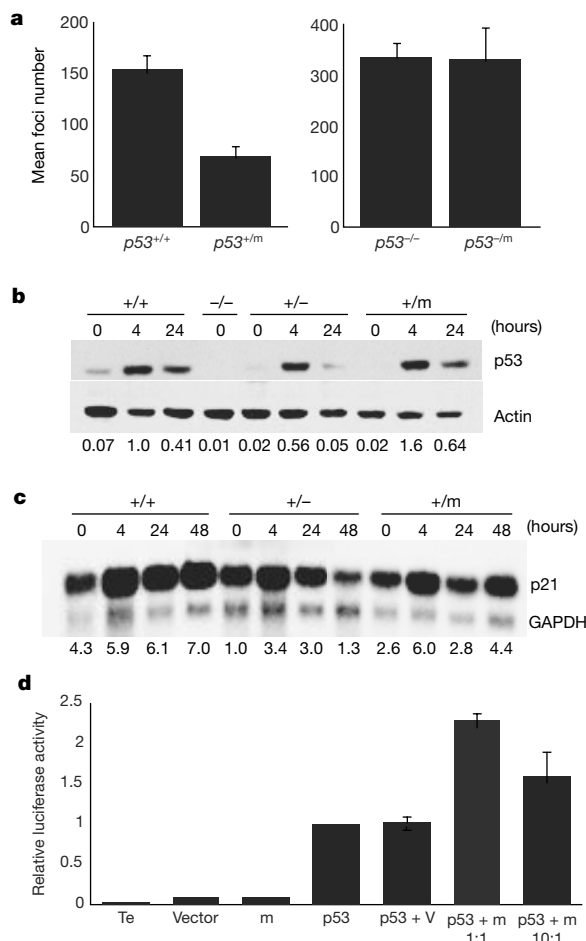
The hair on the older  $p53^{+/m}$  mice appeared to be sparser than that of their  $p53^{+/+}$  counterparts. Hair sparseness could be a function of the ratio of hair follicle cells in the anagen (growth) phase to those in the telogen (resting) phase. Using a hair-growth assay—which involves shaving a dorsal segment of skin on the mouse and measuring the amount of hair growth after 20–25 days—hair growth declines linearly as a function of age in mice<sup>27</sup>. Only a modest reduction in hair growth was observed in the young  $p53^{+/m}$  mice compared with young  $p53^{+/+}$  mice (Fig. 6h). In 22–26-month-old  $p53^{+/m}$  mice, almost no hair growth was observed, whereas old  $p53^{+/+}$  mice displayed robust hair growth (Fig. 6g, h).

A reduced ability to tolerate stresses is a hallmark of ageing<sup>28</sup>. One such age-related stress, wound healing, is delayed as a function of age in both mice and humans<sup>14,29</sup>. To test this capacity, we placed two 3-mm punch biopsies in the dorsal skin of young and old anaesthetized  $p53^{+/m}$  and  $p53^{+/+}$  mice. However, we found that many of the old  $p53^{+/m}$  mice died after injection with the standard dosage of the anaesthetic Avertin, and a greatly reduced dose had to be used for this age/genotype category. This observation agrees with reports that aged organisms are less able to tolerate the stress of anaesthesia than are younger organisms<sup>30,31</sup>.

After adjustment of anaesthetic doses, wounds were induced and monitored over a period of 4 days for closure before sacrifice. Wound diameters at 2, 3 and 4 days after wounding were similar for wounds of young  $p53^{+/m}$  and  $p53^{+/+}$  mice (Fig. 7a); however, old  $p53^{+/m}$  mice exhibited a significant delay in wound closure (Fig. 7b). When cross-sections of the wound at 4 days after wounding were examined by histopathology, there was a frequent reduction in re-epithelialization from the wound edge in the old  $p53^{+/m}$  wound compared with that in the wounds of old  $p53^{+/+}$  mice (Fig. 6c, d).

An acute stress that measures the response to haematopoietic precursor cell ablation is treatment with 5-fluorouracil (5-FU)<sup>32</sup>. In this assay, we found that young  $p53^{+/+}$  and  $p53^{+/m}$  mice showed reduced white blood cell counts at day 6 after 5-FU injection, but had recovered by day 11 (Fig. 7e). The old  $p53^{+/+}$  mice were not overtly affected by 5-FU treatment, and their white blood cell counts had recovered by day 11. However, the  $p53^{+/m}$  mice displayed a marked morbidity after treatment, and one out of six of the old  $p53^{+/m}$  mice injected with 5-FU died. Although old  $p53^{+/+}$  and  $p53^{+/m}$  mice showed similar reductions in white blood cell counts at 6 days after 5-FU injection, the recovery at day 11 was significantly more robust for  $p53^{+/+}$  mice than for  $p53^{+/m}$  mice ( $P = 0.02$ ) (Fig. 7f).

Although a number of age-associated phenotypes appeared earlier in the  $p53^{+/m}$  mice than in the wild-type mice, other age-related changes were not noted. Examination of lungs, heart, brain and intestines failed to reveal any clear differences in the old  $p53^{+/+}$  and  $p53^{+/m}$  mice. Ageing phenotypes observed in other models, such as liver pathologies<sup>20</sup>, hair greying and alopecia<sup>18</sup>, atrophy of intestinal villi<sup>18</sup>, skin ulceration<sup>18,33</sup>, amyloid deposits<sup>33</sup>, brain atrophy<sup>33</sup>, joint



**Figure 3** Transformation and p53 response phenotypes. **a**, Transformation susceptibility of  $p53^{+/+}$ ,  $p53^{+/m}$ ,  $p53^{-/-}$  and  $p53^{-/-m}$  embryonic fibroblasts. Left, Comparison of total foci produced by *ras* plus *myc* retroviral infection of 4 plates each of  $p53^{+/m}$  and  $p53^{+/+}$  MEFs. Right, comparison of total foci produced by *ras* plus *myc* retroviral infection of  $p53^{-/-}$  and  $p53^{-/-m}$  MEFs. **b**, Analysis of p53 protein in kidney cells of  $p53^{+/+}$ ,  $p53^{+/m}$  and  $p53^{-/-}$  mice before and after treatment with 5 Gy of ionizing radiation. The top panel shows p53 levels detected with an antibody to p53; the lower panel displays actin levels. Values below each lane indicate relative p53 levels after normalization to actin. **c**, Northern blot hybridization showing expression of p21<sup>Waf1/Cip1</sup> mRNA in MEFs of  $p53^{+/+}$ ,  $p53^{+/m}$  and  $p53^{-/-}$  mice before and after 5 Gy of ionizing radiation. The upper band is the p21 mRNA signal; the lower band contains the GAPDH signal. Values below each lane indicate relative p21 mRNA levels after normalization to GAPDH levels. **d**, Enhancement of wild-type p53 transactivation activity by the m allele. Saos-2 cell co-transfection assays were performed with a p21 promoter-driven luciferase construct and various combinations of wild-type p53, m allele, or empty vector expression constructs. Relative luciferase activity was normalized to a value of 1.0 produced by wild-type p53 alone. 1:1 and 10:1 refer to the ratio (in  $\mu$ g) of wild-type p53 to m-allele expression construct for each co-transfection. Te, mock transfected cells; V, empty vector.

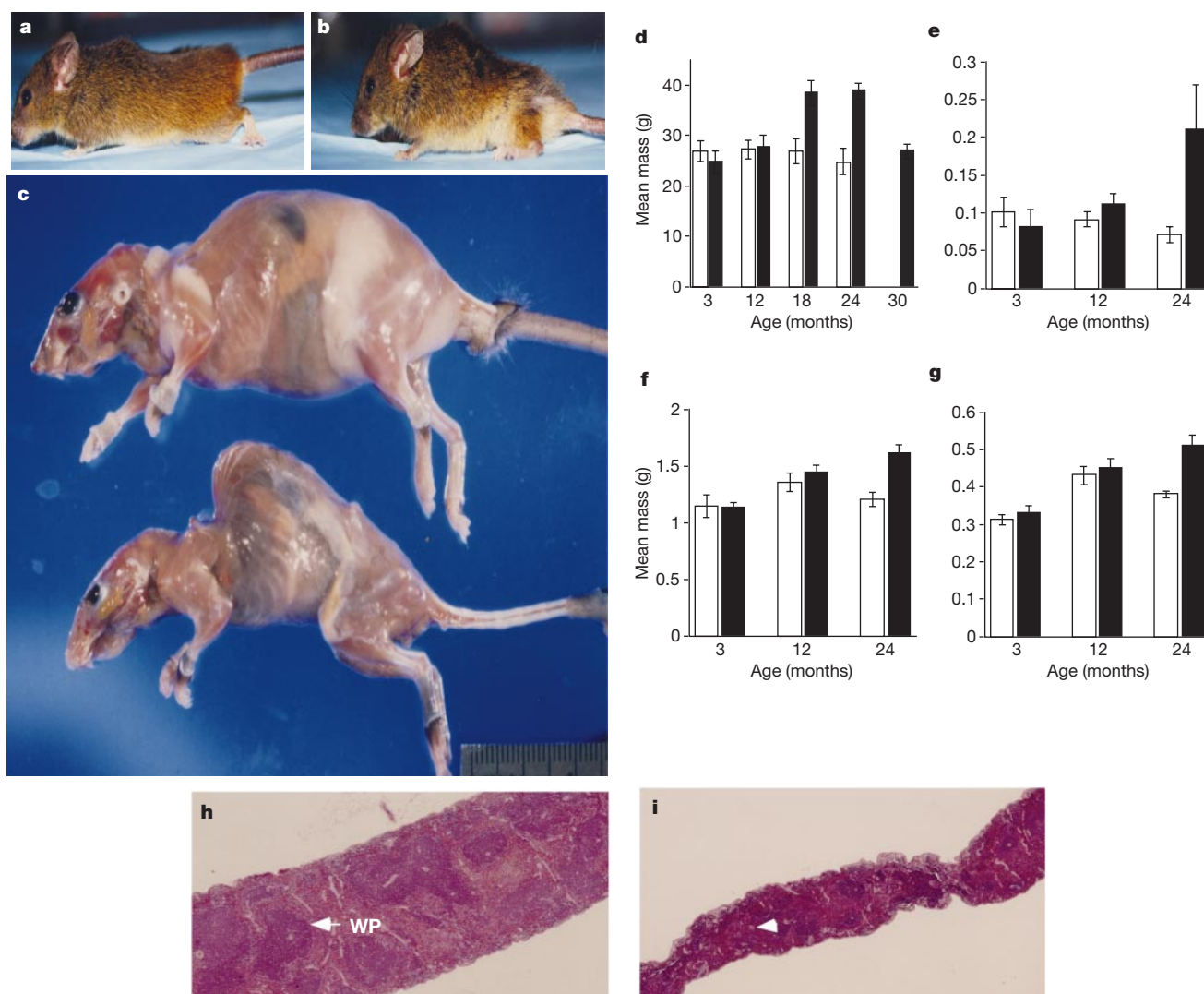


diseases<sup>33</sup>, or cataracts<sup>33</sup> were not increased in frequency in the older  $p53^{+/m}$  mice. Blood chemistry, blood cell counts and male fecundity in the older  $p53^{+/m}$  mice did not differ from that of older  $p53^{+/+}$  mice (Table 2). Thus, the older  $p53^{+/m}$  mice display an early onset of some, but not all, phenotypes associated with ageing.

### pL53 mice exhibit early ageing-related phenotypes

During our observations on the ageing  $p53^{+/m}$  mice, we monitored a colony of mutant p53 transgenic (pL53) mice containing roughly 20 copies of a mutation at codon 135 (Ala135Val), which has been described previously<sup>22</sup>. The protein produced by this mutant allele of p53 is temperature sensitive, and assumes a wild-type conformation at 32.5 °C and a mutant conformation at 37.5 °C (ref. 34). This mutant p53 transgenic mouse is modestly susceptible to a broad array of tumour types, with a 20% tumour incidence by 18 months of age (Table 1)<sup>22</sup>. However, most of the pL53 transgenic mice remain tumour free at 18 months, yet exhibit a generalized lack of robustness compared with their non-transgenic littermates (Fig. 8a, b). Some phenotypes exhibited by the pL53 transgenic mice were

similar to those of the older  $p53^{+/m}$  mice. For example, by 18–20 months, many of the pL53 mice exhibited sparse, ruffled fur, loss of weight, lordokyphosis and lethargy. Necropsies of these mice usually failed to reveal any overt or microscopic tumours or other pathologies. However, X-rays of representative 18-month-old animals showed evidence of increased osteoporosis in pL53 transgenic mice (Fig. 8c). The ability of the pL53 transgenic mice to regrow hair was also markedly diminished in comparison with matched non-transgenic littermates (Fig. 8d). Moreover, cross-sectional analysis of skin indicated that the transgenic mice had significantly reduced subcutaneous adipose tissue compared with their age- and sex-matched normal littermates ( $P = 0.001$ ) (Fig. 8e). Finally, the wound-healing assay showed a significant reduction in wound closure for the pL53 transgenic mice 4 days after induction of a 3-mm punch biopsy in the dorsal skin ( $P = 0.05$ ) (Fig. 8f). Although body weight was often reduced in the 18-month-old pL53 transgenic mice in comparison with similarly aged non-transgenic mice, organ weights (spleen, kidney, liver, testes, skeletal muscle) were only modestly affected. Thus, the older



**Figure 4** Phenotypes of aged  $p53^{+/+}$  and  $p53^{+/m}$  mice. **a**, Photograph of a 24-month-old  $p53^{+/+}$  male mouse. **b**, Photograph of a 24-month-old  $p53^{+/m}$  male mouse showing reduced size and lordokyphosis. **c**, Representative 20-month-old skinned  $p53^{+/+}$  female mouse (top) and age- and sex-matched  $p53^{+/m}$  mouse (bottom). Pronounced lordokyphosis, loss of body mass, muscle atrophy and loss of adipose tissue is evident in the  $p53^{+/m}$  mouse. **d–g**, Comparisons of mean  $p53^{+/+}$  (black columns) and  $p53^{+/m}$

(white columns) body (**d**), spleen (**e**), liver (**f**) and kidney (**g**) masses. **h**, **i**, Haematoxylin and eosin-stained cross section of 24-month-old  $p53^{+/+}$  spleen (**h**) and  $p53^{+/m}$  spleen (**i**) at  $\times 10$  magnification. White pulp (WP) is indicated by an arrow in the  $p53^{+/+}$  spleen. The lack of arteriole-associated white pulp is indicated by an arrow in the  $p53^{+/m}$  spleen.

pL53 transgenic mice recapitulated some of the ageing-related phenotypes observed in the  $p53^{+/m}$  mice, particularly those related to skin and bone.

## Discussion

The  $p53^{+/m}$  mice generated in this study were highly resistant to spontaneous tumours. This resistance depends on wild-type p53, as in its absence (for example, the  $p53^{-/-}$  mice) the m allele confers little protection from tumours. This is consistent with our hypothesis that the gene product of the m allele augments the activity of wild-type p53 to prevent tumour formation. The tumour resistance of  $p53^{+/m}$  mice may be partly intrinsic to the individual cells of the animals, as fibroblasts derived from  $p53^{+/m}$  embryos are more resistant to transformation by activated *ras* plus *myc* onco-

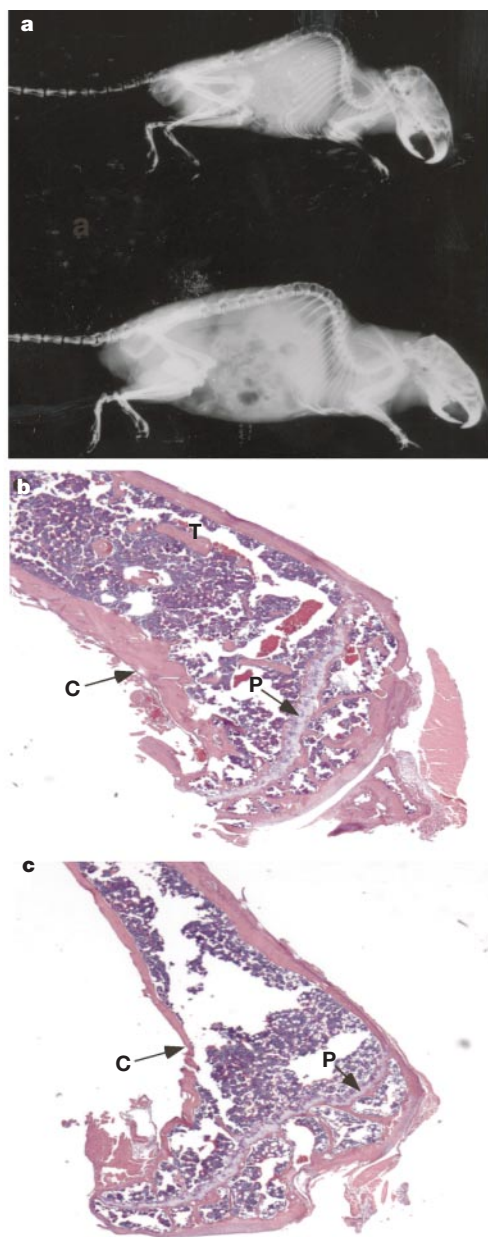
genes. However, this does not rule out the possibility that systemic changes, such as alterations in hormone or growth factor levels, could contribute to the tumour resistance phenotypes. Despite an enhanced tumour resistance, the  $p53^{+/m}$  mice showed a reduced longevity in comparison with their  $p53^{+/+}$  littermates. This reduced longevity is associated with the early appearance of a number of ageing-associated phenotypes. These particular ageing phenotypes correlate with a 23% reduction in median lifespan, suggesting that they are important for organismal longevity.

The extent of the m-allele deletion in the  $p53^{+/m}$  mouse has not been completely defined. The p53-associated deletion includes p53 exons 1–6 as well as sequences upstream of the p53 gene. It is possible that the  $p53^{+/m}$ -associated phenotypes could be a result of either haploinsufficiency in genes upstream of p53 or the production of a truncated p53 gene product. Although haploinsufficiency cannot be ruled out, we favour the hypothesis that a truncated C-terminal gene product of p53 is the primary effector (see below). A truncated message capable of expressing the C-terminal p53 protein is expressed in many tissues of the  $p53^{+/m}$  mouse, and several laboratories have shown that p53 C-terminal fragments or peptides can augment wild-type p53 activities<sup>35–39</sup>. We have shown that p53 m-allele expression constructs enhance wild-type p53 transactivation activity and can suppress cancer cell growth in the presence of wild-type p53 (S.V., S.D.T., N.G. and H.I., unpublished data). The cells of  $p53^{+/m}$  mice have increased wild-type p53 responses to ionizing radiation compared with cells of  $p53^{+/-}$  mice. The enhanced tumour resistance in  $p53^{+/m}$  mice is dependent on wild-type p53, supporting a model in which the m-allele product requires wild-type p53 to promote tumour suppression. Another p53 mutant line (pL53) recapitulates some early ageing phenotypes of  $p53^{+/m}$  mice.

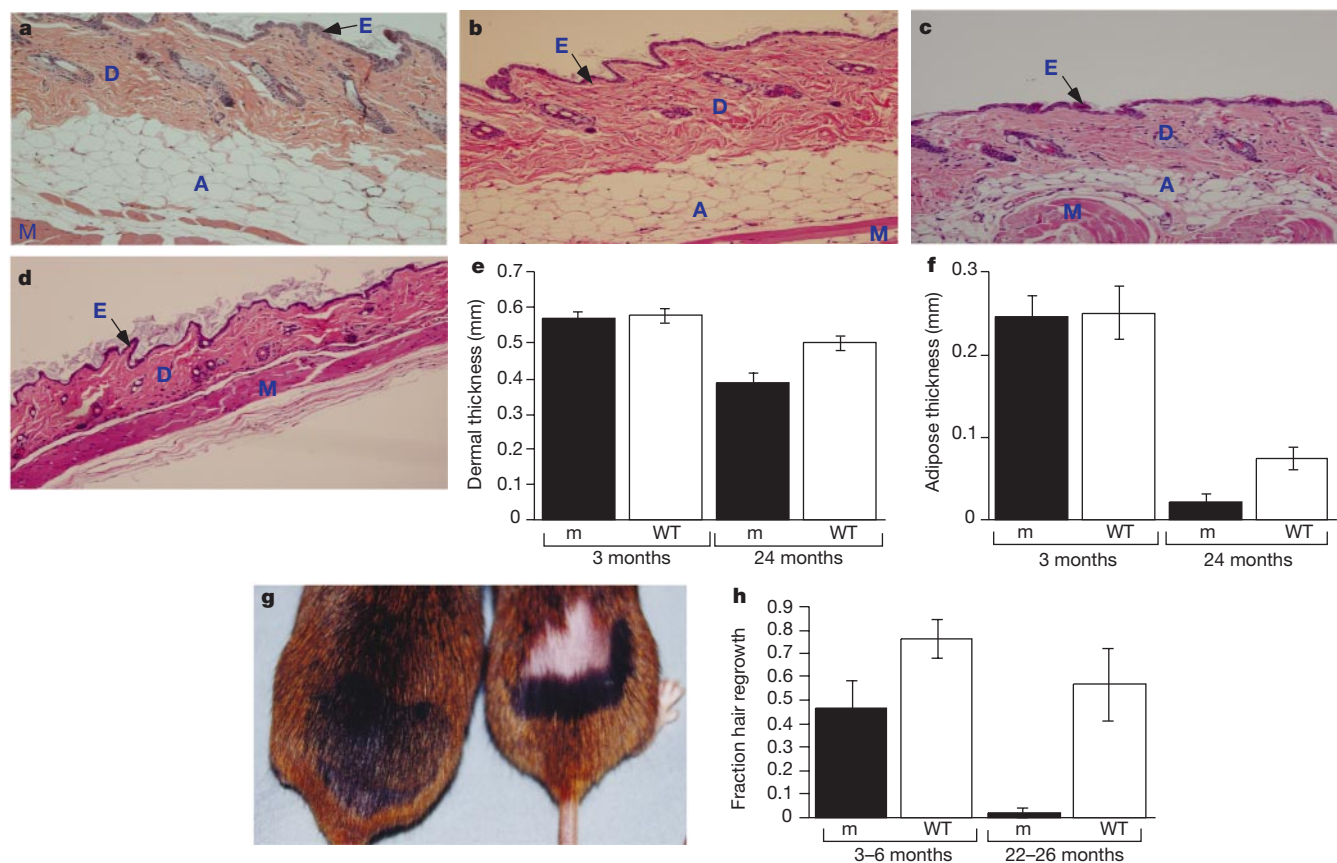
The display of ageing-related phenotypes in the pL53 transgenic mice may seem paradoxical, given that this p53 transgene mutation would be expected to confer a dominant negative activity and inhibit wild-type p53 function. The modest increase in tumour susceptibility indicates that such activity is present. However, because this mutant p53 allele encodes a temperature-sensitive p53 protein, we propose that the pool of transgene-encoded mutant p53 may exist in an equilibrium between wild-type active and mutant inactive conformations. Where body temperatures are presumably lower (near the skin surface), more of the transgene protein may be in a wild-type conformation and thus inhibit proliferation. The reduced adipose tissue, hair growth and wound healing in the older pL53 transgenic animals are consistent with this.

For both  $p53^{+/m}$  and pL53 transgenic models, we propose that the early ageing phenotypes are in part a result of enhanced activity of wild-type p53 in some tissues. The reduced cellularity and mass in organs of the older  $p53^{+/m}$  mice suggests that organ cell numbers are not maintained. Moreover, some of the ageing phenotypes suggest a reduction in proliferation of stem cells. With the ageing process, this proliferative reserve may decline more rapidly in the  $p53^{+/m}$  mice as their stem cells undergo replicative senescence sooner than their  $p53^{+/+}$  counterparts. The accumulation of genetic insults in the stem cells of  $p53^{+/m}$  mice may provoke enhanced arrest responses that gradually result in fewer division-competent cells. The  $p53^{+/m}$  mice eventually reach a point in which the proliferative capacity of stem cells is so reduced that sufficient numbers of mature cells cannot be provided to maintain organ homeostasis. The resulting phenotypes may include reductions in organ mass, function and tolerance for stress.

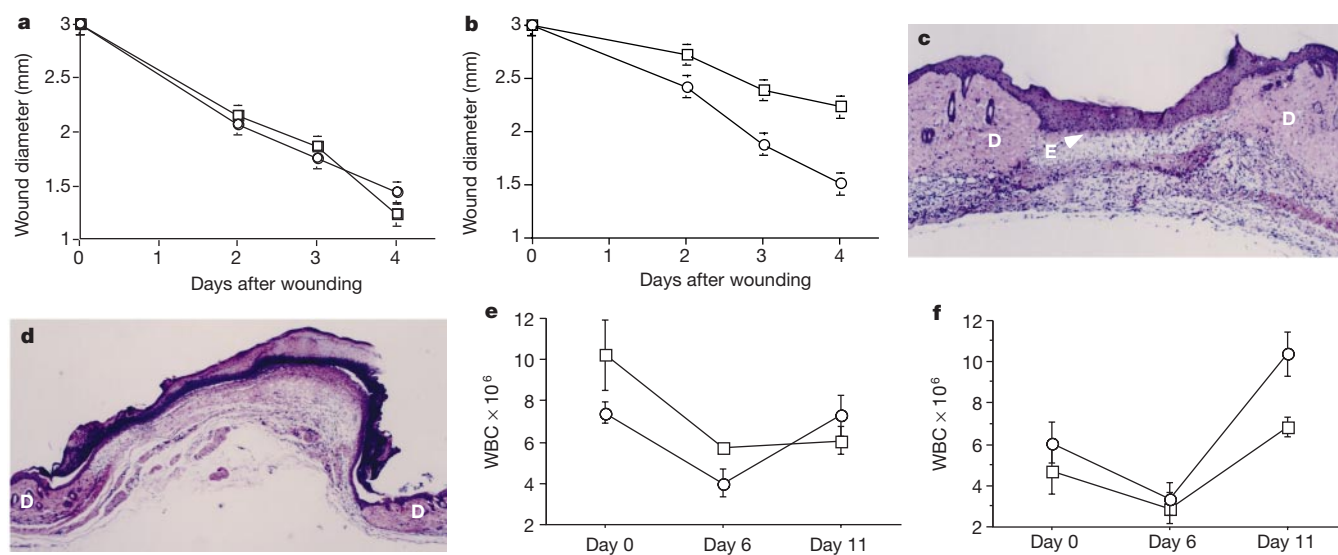
The data presented here support a role for p53 in the regulation of ageing and longevity in mice. The association of early ageing and tumour resistance in the  $p53^{+/m}$  mice is also consistent with the idea that senescence is a mechanism of tumour suppression<sup>4,12,13</sup>. We propose that an ageing-related reduction in stem cell proliferation may have a more important role in longevity than previously recognized. □



**Figure 5** Bone phenotypes in  $p53^{+/+}$  and  $p53^{+/m}$  mice. **a**, Whole-body radiograph of 24-month-old  $p53^{+/m}$  (top) and  $p53^{+/+}$  (bottom) mice. **b**, **c**, Haematoxylin and eosin-stained cross-sections of tibias from a 24-month-old  $p53^{+/+}$  mouse (**b**) and an age-matched  $p53^{+/m}$  mouse (**c**). Note the reduction in cortical bone thickness (C) and absence of trabecular bone (T) in the tibia of the 24-month-old  $p53^{+/m}$  mouse (**c**). P, epiphyseal plate.

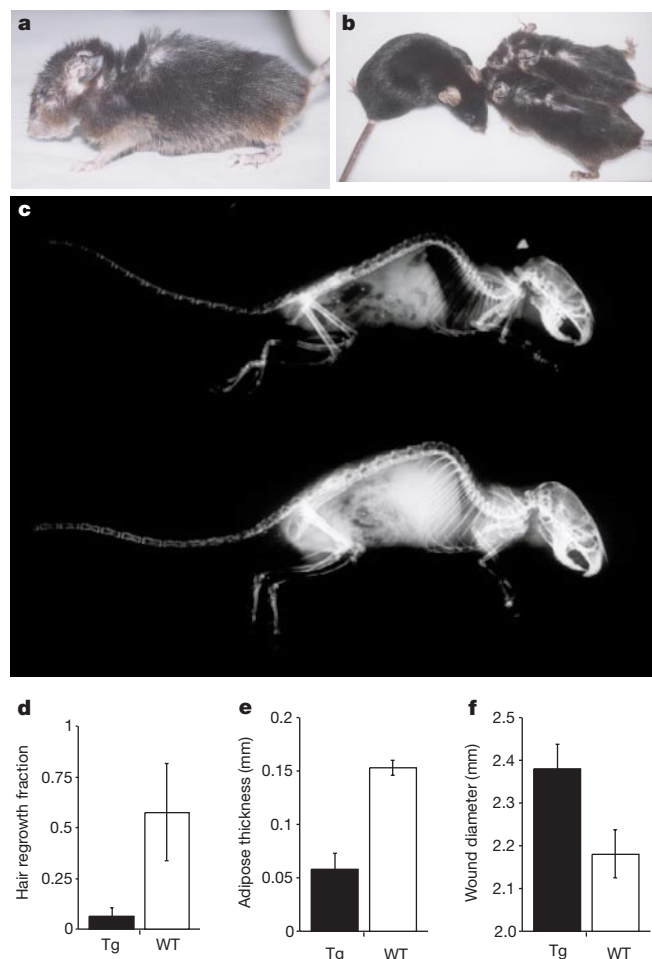


**Figure 6** Skin and hair growth phenotypes in  $p53^{+/+}$  and  $p53^{+/m}$  mice. **a–d**, Cross-sections of dorsal skin from a 3-month-old  $p53^{+/+}$  mouse (**a**), a 3-month-old  $p53^{+/m}$  mouse (**b**), a 24-month-old  $p53^{+/+}$  mouse (**c**) and a 24-month-old  $p53^{+/m}$  mouse (**d**). Note the reduced thickness of the dermis (D) and subcutaneous adipose tissue (A) in the skin of 24-month-old  $p53^{+/m}$  mice (**d**). E, epidermis; M, muscle. Magnification 100 $\times$ . **e, f**, Graphs showing mean dermal thickness (**e**) and mean thickness of subcutaneous adipose layer (**f**) of 3- and 24-month-old  $p53^{+/+}$  and  $p53^{+/m}$  mice. WT, wild type. **g**, Representative 24-month-old  $p53^{+/+}$  mouse (left) and age-matched  $p53^{+/m}$  mouse (right) 20 days after removal of hair from a 2-cm<sup>2</sup> dorsal area. **h**, Mean hair growth in 3- and 24-month-old  $p53^{+/+}$  and  $p53^{+/m}$  mice.



**Figure 7** Stress responses in  $p53^{+/+}$  and  $p53^{+/m}$  mice. **a, b**, Graphs showing mean wound diameter in 3- (**a**) and 24-month-old (**b**)  $p53^{+/+}$  (circles) and  $p53^{+/m}$  (squares) mice at 0, 2, 3 and 4 days after wounding. **c, d**, Wound cross-sections stained with haematoxylin and eosin 4 days after wound induction. Representative wounds from 24-month-old  $p53^{+/+}$  (**c**) and  $p53^{+/m}$  (**d**) mice show reduced re-epithelialization in the  $p53^{+/m}$  mice. D, dermis; E, epithelialization. Magnification 100 $\times$ . **e, f**, White blood cell (WBC) counts in 3-month-old (**e**) and 24-month-old (**f**)  $p53^{+/+}$  (circles) and  $p53^{+/m}$  (squares) mice at 0, 6 and 11 days after 5-FU injection.





**Figure 8** Ageing-related phenotypes observed in pL53 transgenic mice. **a**, Photograph of 20-month-old pL53 transgenic male mouse. **b**, Photograph of two pL53 transgenic (right) and one non-transgenic (left) littermates at 20 months of age. The bedraggled appearance of the transgenic mice was not due to fighting as each of these mice were caged individually. Moreover, all three mice were devoid of tumours. **c**, Radiographs of 18-month-old female pL53 transgenic mouse (top) and non-transgenic mouse (bottom). Note the reduction in bone density in the transgenic female. **d**, Mean hair growth in 16–18-month-old pL53 (Tg) and non-transgenic (WT) mice. **e**, Mean subcutaneous adipose thickness in pL53 transgenic and wild-type 16–18-month-old mice. **f**, Mean wound diameter 4 days after induction by 3-mm biopsy in 16–18-month-old wild-type and pL53 transgenic mice.

## Methods

### Generation of $p53^{+/m}$ mice

The 'hit-and-run' gene targeting approach<sup>40</sup> was used to introduce a mutant p53 targeting construct derived from an exon 2 (*XhoI*) to exon 10 (*SacI*) genomic clone containing a Arg245Trp point mutation and linked to a PGK promoter-driven hypoxanthine phosphoribosyltransferase (HPRT) selection cassette. Along with the point mutation, we incorporated a *StuI* site to facilitate subsequent genotyping efforts. One targeted embryonic stem cell clone was isolated that showed incorporation of the point mutation into the endogenous p53 allele but exhibited a deletion of p53 exons 1–6. This clone, containing the p53 mutation designated 'm', was amplified and used to generate chimaeric and germline  $p53^{+/m}$  mice according to standard methods<sup>41</sup>.

### Mouse breeding, maintenance and longevity

We generated the mutant p53 m allele in AB1 embryonic stem cells of 129/Sv origin. Chimaeric mice with this mutation were backcrossed two generations into C57BL/6, so that the mice in the  $p53^{+/m}$  study were mixed inbred C57BL/6–129/Sv mice.  $p53^{+/m}$  mice of this background were intercrossed with  $p53^{+/-}$  mice of similar background<sup>19</sup>, and the  $p53^{+/+}$ ,  $p53^{+/m}$ ,  $p53^{+/-}$  and  $p53^{-/-}$  offspring were monitored for tumour formation and longevity.

The pL53 transgenic mice<sup>22</sup> were provided by Alan Bernstein in a CD-1 background, and these mice were backcrossed twice into a C57BL/6 background before intercrossing. pL53 transgenic mice were always crossed to  $p53^{+/+}$  mice of the same background to maintain transgene hemizygosity in the offspring. Littermate mice of different p53 genotypes were usually housed together and fed freely with standard mouse chow over their lifespan. Although the mice were not housed in a pathogen-free environment, morbidity owing to infectious agents was not noted in the colony. Mice exhibiting overt tumours or extreme morbidity were killed and subjected to necropsy.

### Mouse genotyping

Two Southern blot assays were necessary to distinguish  $p53^{+/m}$ ,  $p53^{+/+}$ ,  $p53^{+/-}$ ,  $p53^{-/-}$  and  $p53^{-/m}$  mice. In the first assay to distinguish the '−' and '+' alleles, tail DNAs were cleaved with *Bam*HI and hybridized to a p53 exon 2–6 cDNA probe as described previously<sup>19</sup>. In the second assay to distinguish the m allele, tail DNAs were cleaved with *Stu*I and hybridized to a p53 exon 5–10 probe. A fragment migrating at 0.9 kb was diagnostic for the m allele. To genotype pL53 transgenic mice, we hybridized tail DNAs cleaved with *Bam*HI with the p53 exon 2–6 cDNA probe. Mice with the mutant p53 transgene exhibit a 10–20-fold increase in hybridization intensity of a 5-kb, p53-specific fragment compared with their non-transgenic counterparts.

### RNA analyses

Total RNA was purified from frozen tissue using Trizol Reagent (Gibco BRL) according to the manufacturer's specifications. RT-PCR specific for the m-allele messages and p53 messages was performed on 1 µg total RNA. The m-allele message and p53 message were amplified using PCR primers specific for the leader sequence of the m allele (5'-CGGGCAGGTTCTGAAGGCAGTTTC-3') and exon 7 of the wild-type p53 allele (5'-TGCATGGGGGCATGAACCGCCGACC-3'). The downstream primer (5'-CAGCTCCCGGAACATCTCGAA-3') was anchored in p53 exon 10. 5' RACE was performed using the Marathon 5' RACE Kit (Clontech). One microgram of total RNA from spleens of  $p53^{+/+}$  and  $p53^{+/m}$  mice was used to distinguish the m-allele message. We gel isolated, cloned and sequenced the differentially migrating band in the  $p53^{+/m}$  sample. We performed northern and western blot assays as described previously<sup>42</sup>.

### Transformation and transfection assays

Primary MEFs were plated on 100-mm plates at a density of  $10^6$  cells per plate in 15% fetal calf serum (FCS) and DMEM growth medium. Cells were treated 24 h after plating with varying titres of a *ras/myc* retrovirus diluted in DMEM without FCS for 2 h. After treatment, we added 15% FCS DMEM to the plates. We replaced growth medium every 3 days for 3 weeks. Transformed foci were examined microscopically for transformed cell characteristics before counting. We counted the foci after Giemsa staining. Representative foci of  $p53^{+/m}$  and  $p53^{+/+}$  mice were amplified, and  $10^6$  cells were injected subcutaneously into nude mice. Virtually all foci of both genotypes produced tumours in nude mice. To assay p53 transactivation activity *in vitro*, 60-mm plates of subconfluent Saos-2 cells were transfected with 1 µg of a p21 promoter-luciferase plasmid and 2 µg each of various combinations of cytomegalovirus (CMV) promoter-driven wild-type p53, m allele and empty vector expression plasmids using Lipofectamine Plus reagent (Invitrogen). Cell lysates were prepared 48 h after transfection, and luciferase activity was measured using a Turner TD-20e luminometer and normalized to renilla luciferase according to the instructions provided with the dual-luciferase assay kit (Promega).

### Bone analyses

Mice were X-rayed *in situ* under anaesthesia for whole-body X-rays. We used an X-ray dose of 20 kV for 20 s. We prepared and analysed cross-sections of tibias as described previously<sup>43</sup>.

### Skin, wound healing and hair growth assays

Dorsal skin sections were fixed, embedded in paraffin and stained with haematoxylin and eosin. The thickness of the dermal and adipose layers from the skin samples were determined by taking 10 random measurements along the length of individual skin samples ( $n = 4$ ) for each genotype and age group. Wound healing was assayed as described previously<sup>44</sup>, except that full thickness wounds were generated with a 3-mm biopsy punch rather than a 5-mm biopsy punch. For the hair growth assay, age-matched  $p53^{+/+}$  and  $p53^{+/m}$  mice (3–6 months and 20–26 months) were shaved on their dorsal surface using an electric razor. We measured hair regrowth 20 days after shaving as described previously<sup>27</sup>.

### Haematopoietic cell ablation with 5-fluorouracil

Two-month-old and 20-month-old  $p53^{+/+}$  and  $p53^{+/m}$  mice were treated with an intraperitoneal bolus injection of 5-FU (100 mg per kg body weight). Peripheral blood was obtained by tail bleeds before injection, and on days 6 and 11 after injection. We analysed peripheral blood counts and haemoglobin amounts by flow cytometry (Technicon Instruments).

Received 29 June; accepted 24 October 2001.

- Levine, A. J. p53, the cellular gatekeeper for growth and division. *Cell* **88**, 323–331 (1997).
- Ko, L. J. & Prives, C. p53: puzzle and paradigm. *Genes Dev.* **10**, 1054–1072 (1996).
- Giaccia, A. J. & Kastan, M. B. The complexity of p53 modulation: emerging patterns from divergent signals. *Genes Dev.* **12**, 2973–2983 (1998).
- Itahana, K., Dimri, G. & Campisi, J. Regulation of cellular senescence by p53. *Eur. J. Biochem.* **268**, 2784–2791 (2001).

5. Atadja, P., Wong, H., Garkavtsev, I., Geillette, C. & Riabowol, K. Increased activity of p53 in senescing fibroblasts. *Proc. Natl Acad. Sci. USA* **92**, 8348–8352 (1995).
6. Bond, J. A. *et al.* Evidence that transcriptional activation by p53 plays a direct role in the induction of cellular senescence. *Oncogene* **13**, 2097–2104 (1996).
7. Webley, K. *et al.* Posttranslational modifications of p53 in replicative senescence overlapping but distinct from those induced by DNA damage. *Mol. Cell. Biol.* **20**, 2803–2808 (2000).
8. Shay, J. W., Pereira-Smith, O. M. & Wright, W. E. A role for both RB and p53 in the regulation of human cellular senescence. *Exp. Cell Res.* **196**, 33–39 (1991).
9. Bond, J. A., Wyllie, F. S. & Wynford-Thomas, D. Escape from senescence in human diploid fibroblasts induced directly by mutant p53. *Oncogene* **9**, 1885–1889 (1994).
10. Gire, V. & Wynford-Thomas, D. Reinitiation of DNA synthesis and cell division in senescent human fibroblasts by microinjection of anti-p53 antibodies. *Mol. Cell. Biol.* **18**, 1611–1621 (1998).
11. Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D. & Lowe, S. W. Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16<sup>INK4a</sup>. *Cell* **88**, 593–602 (1997).
12. Sager, R. Senescence as a mode of tumor suppression. *Environ. Health Perspect.* **93**, 59–62 (1991).
13. Campisi, J. Aging and cancer: the double-edged sword of replicative senescence. *J. Am. Geriatric Soc.* **45**, 482–488 (1997).
14. Rudolph, K. L. *et al.* Longevity, stress response, and cancer in aging telomerase-deficient mice. *Cell* **96**, 701–712 (1999).
15. Chin, L. *et al.* p53 deficiency rescues the adverse effects of telomere loss and cooperates with telomere dysfunction to accelerate carcinogenesis. *Cell* **97**, 527–538 (1999).
16. Vogel, H., Lim, D. S., Karsenty, G., Finegold, M. & Hasty, P. Deletion of Ku86 causes early onset of senescence in mice. *Proc. Natl Acad. Sci. USA* **96**, 10770–10775 (1999).
17. Lim, D. S. *et al.* Analysis of ku80-mutant mice and cells with deficient levels of p53. *Mol. Cell. Biol.* **20**, 3772–3780 (2000).
18. Migliaccio, E. *et al.* The p66<sup>shc</sup> adaptor protein controls oxidative stress response and life span in mammals. *Nature* **402**, 309–313 (1999).
19. Donehower, L. A. *et al.* Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature* **356**, 215–221 (1992).
20. Jacks, T. *et al.* Tumor spectrum analysis in p53-mutant mice. *Curr. Biol.* **4**, 1–7 (1994).
21. Purdie, C. A. *et al.* Tumour incidence, spectrum and ploidy in mice with a large deletion in the p53 gene. *Oncogene* **9**, 603–609 (1994).
22. Lavigne, A. *et al.* High incidence of lung, bone, and lymphoid tumors in transgenic mice overexpressing mutant alleles of the p53 oncogene. *Mol. Cell. Biol.* **9**, 3982–3991 (1989).
23. Arking, R. *Biology of Aging* 2nd edn 153–250 (Sinauer, Sunderland, Massachusetts, 1998).
24. Kalu, D. N. in *Handbook of Physiology, Section 11: Aging* (ed. Masoro, E. J.) 395–412 (Oxford Univ. Press, New York, 1995).
25. Weiss, A., Arbell, I., Steinhagen Thiessen, E. & Silberman, M. Structural changes in aging bone: osteopenia in the proximal femurs of female mice. *Bone* **12**, 165–172 (1991).
26. Chuttani, A. & Gilchrist, B. A. in *Handbook of Physiology, Section 11: Aging* (ed. Masoro, E. J.) 309–324 (Oxford Univ. Press, New York, 1995).
27. Harrison, D. E. & Archer, J. R. Biomarkers of aging: tissue markers. Future research needs, strategies, directions and priorities. *Exp. Gerontol.* **23**, 309–321 (1988).
28. Shock, N. W. Aging of physiological systems. *J. Chronic Dis.* **36**, 137–142 (1983).
29. Gerstein, A. D., Phillips, T. J., Rogers, G. S. & Gilchrist, B. A. Wound healing and aging. *Dermatol. Clin.* **11**, 749–757 (1993).
30. Muravchik, S. in *Clinical Anesthesia* 3rd edn (eds Barash, P. G., Cullen, B. F. & Stoelting, R. K.) 1125–1136 (Lippincott-Raven, Philadelphia, 1997).
31. Harvey, R. C. & Paddleford, R. R. Management of geriatric patients. A common occurrence. *Vet. Clin. North Am. Small Anim. Pract.* **29**, 683–699 (1999).
32. Harrison, D. E. Evaluating functional abilities of primitive hematopoietic stem cell populations. *Curr. Top. Microbiol. Immunol.* **177**, 13–30 (1992).
33. Takeda, T. *et al.* Pathobiology of the senescence-accelerated mouse (SAM). *Exp. Gerontol.* **32**, 117–127 (1997).
34. Michalovitz, D., Halevy, O. & Oren, M. Conditional inhibition of transformation and of cell proliferation by a temperature-sensitive mutant of p53. *Cell* **62**, 671–680 (1990).
35. Hupp, T. R., Sparks, A. & Lane, D. P. Small peptides activate the latent sequence-specific DNA binding function of p53. *Cell* **83**, 237–245 (1995).
36. Jayaraman, J. & Prives, C. Activation of p53 sequence-specific DNA binding by short single strands of DNA requires the p53 C-terminus. *Cell* **81**, 1021–1029 (1995).
37. Muller-Tiemann, B. F., Halazonetis, T. D. & Elting, J. J. Identification of an additional negative regulatory region for p53 sequence-specific DNA binding. *Proc. Natl Acad. Sci. USA* **95**, 6079–6084 (1998).
38. Selivanova, G. *et al.* Restoration of the growth suppression function of mutant p53 by a synthetic peptide derived from the p53 C-terminal domain. *Nature Med.* **3**, 632–638 (1997).
39. Selivanova, G., Rybachenko, L., Jansson, E., Iotsova, V. & Wiman, K. G. Reactivation of mutant through interaction of a C-terminal peptide with the core domain. *Mol. Cell. Biol.* **19**, 3395–3402 (1999).
40. Hasty, P., Ramirez-Solis, R., Krumlauf, R. & Bradley, A. Introduction of a subtle mutation into the Hox-2.6 locus in embryonic stem cells. *Nature* **350**, 243–246 (1991).
41. Hogan, B., Beddington, R., Costantini, F. & Lacy, E. *Manipulating the Mouse Embryo: A Laboratory Manual* 2nd edn 189–216 (Cold Spring Harbor Laboratory Press, New York, 1994).
42. Venkatachalam, S. *et al.* Retention of wild-type p53 in tumors from p53 heterozygous mice: reduction of p53 dosage can promote cancer formation. *EMBO J.* **17**, 4657–4667 (1998).
43. Ducy, P. *et al.* Increased bone formation in osteocalcin-deficient mice. *Nature* **382**, 448–452 (1996).
44. Wojcik, S. M., Bundman, D. S. & Roop, D. R. Delayed wound healing in keratin 6a knockout mice. *Mol. Cell. Biol.* **20**, 5248–5255 (2000).

## Acknowledgements

We thank X.-J. Wang, G. Van Zant, D. Roop, R. Waikel, P. Biggs, M. Patel, S. Wojcik, R. Levasseur, V. Hortenstine, R. Ford, S. Wojcik, C. Pickering, R. Geske and M. Oren for advice and technical assistance. We also thank G. Lozano for luciferase and p53 plasmids. This study was supported by the National Cancer Institute.

Correspondence and requests for materials should be addressed to L.A.D. (e-mail: larryd@bcm.tmc.edu).

# The cosmological density of baryons from observations of $^3\text{He}^+$ in the Milky Way

T. M. Bania\*, Robert T. Rood† & Dana S. Balser‡

\* Institute for Astrophysical Research, Boston University,

725 Commonwealth Avenue, Boston, Massachusetts 02215, USA

† Astronomy Department, University of Virginia, PO Box 3818, Charlottesville, Virginia 22903-0818, USA

‡ National Radio Astronomy Observatory, PO Box 2, Green Bank, West Virginia 24944, USA

Primordial nucleosynthesis after the Big Bang can be constrained by the abundances of the light elements and isotopes  $^2\text{H}$ ,  $^3\text{He}$ ,  $^4\text{He}$  and  $^7\text{Li}$  (ref. 1). The standard theory of stellar evolution predicts that  $^3\text{He}$  is also produced by solar-type stars<sup>2</sup>, so its abundance is of interest not only for cosmology, but also for understanding stellar evolution and the chemical evolution of the Galaxy. The  $^3\text{He}$  abundance in star-forming (H II) regions agrees with the present value for the local interstellar medium<sup>3</sup>, but seems to be incompatible<sup>4–6</sup> with the stellar production rates inferred from observations of planetary nebulae<sup>7</sup>, which provide a direct test of stellar evolution theory<sup>8</sup>. Here we develop our earlier observations<sup>9,10</sup>, which, when combined with recent theoretical developments in our understanding of light-element synthesis and destruction in stars<sup>11–14</sup>, allow us to determine an upper limit for the primordial abundance of  $^3\text{He}$  relative to hydrogen:  $^3\text{He}/\text{H} = (1.1 \pm 0.2) \times 10^{-5}$ . The primordial density of all baryons determined from the  $^3\text{He}$  data is in excellent agreement with the densities calculated from other cosmological probes. The previous conflict is resolved because most solar-mass stars do not produce enough  $^3\text{He}$  to enrich the interstellar medium significantly.

We obtain the  $^3\text{He}$  abundance in Milky Way sources using measurements of the 8.665 GHz (3.46 cm) spin-flip transition of  $^3\text{He}^+$ . To derive the most accurate possible  $^3\text{He}/\text{H}$  abundances requires the simultaneous development of a theoretical framework that allows us to understand quantitatively the inherent systematic errors. As is the case for any cosmic abundance determination, converting the measured  $^3\text{He}^+$  column density into an abundance ratio relative to hydrogen,  $^3\text{He}/\text{H}$ , is nontrivial. For the case of  $^3\text{He}$  this conversion depends on the density and ionization structure of each nebula. The source sample is currently comprised of 60 H II regions and 6 planetary nebulae. Because the Milky Way interstellar medium is optically thin at centimetre wavelengths, our source sample probes a larger volume of the Galactic disk than does any other light-element tracer of Galactic chemical evolution.

While numerically modelling our  $^3\text{He}^+$  sources we discovered that a major advantage of large, diffuse H II regions is that they tend to be 'simple', that is, the H II region has a relatively simple density and ionization structure. This allows us to derive abundances from the observed  $^3\text{He}^+$  emission line parameters without recourse to the rather complex modelling required for most classic H II regions<sup>10</sup>. Because the abundance corrections for simple sources are small, these sources can in principle yield the most accurate  $^3\text{He}/\text{H}$  abundance ratios attainable. Sources that are 'simple' in the sense just described were especially well suited for observation with the National Radio Astronomy Observatory (NRAO) 140-foot telescope, so during the last few years of its operation we worked to increase the size of our simple source sample. Our  $^3\text{He}$  H II region sample has 21 simple sources for which we can determine abundances (Table 1). The remaining 39 sources are a mixture of complex sources,  $^3\text{He}^+$  detections for which we do not yet have

adequate continuum data to determine the H column depth, and  $^3\text{He}^+$  non-detections.

Because H II regions are objects of zero age when compared to the age of the Galaxy, the elemental abundances of these nebulae chronicle the results of billions of years of Galactic chemical evolution. Measurement of the present  $^3\text{He}$  abundance should therefore be an important diagnostic of chemical evolution in the Galaxy. Standard stellar  $^3\text{He}$  nucleosynthesis predicts that the  $^3\text{He}/\text{H}$  abundance ratio should grow with time and be higher in those parts of the Galaxy where there has been substantial stellar processing. Specifically, (1) there should be a  $^3\text{He}/\text{H}$  abundance gradient across the Galactic disk with the highest abundances occurring in the highly processed inner Galaxy; (2) the  $^3\text{He}/\text{H}$  abundance should grow with source metallicity; and (3) the protosolar  $^3\text{He}/\text{H}$  abundance should be less than that found in the present interstellar medium. None of these predictions is confirmed by observations.

Knowing the contribution of planetary nebulae to the  $^3\text{He}$  abundance in the Milky Way interstellar medium is therefore essential for understanding the chemical evolution of the Galaxy<sup>15</sup>. In our search for  $^3\text{He}^+$  emission from planetary nebulae we were working at the sensitivity limit of the 100-m telescope operated by the Max-Planck-Institut für Radioastronomie (MPIfR). Our survey sample of 6 planetary nebulae was therefore chosen to maximize the likelihood of  $^3\text{He}^+$  detections, and so it was highly biased. Planetary nebulae with high nitrogen or  $^4\text{He}$  abundances were excluded from the sample because either of these would have been a signature of mixing in the stellar interior that would have destroyed  $^3\text{He}$ . The planetary nebulae survey showed that there is some stellar production of  $^3\text{He}$ .

We independently confirmed our result for the planetary nebula NGC3242 using the NRAO 140-foot telescope. After modelling this source we conclude that a  $^3\text{He}/\text{H}$  abundance ratio by number of  $(2-5) \times 10^{-4}$  is consistent with both observed spectra. This

Table 1  $^3\text{He}$  sample of simple Galactic H II regions

| Source    | $R_{\text{gal}}$ (kpc) | [O/H]* | $^4\text{He}^+/\text{H}^+$ | $^3\text{He}^+/\text{H}^+ (10^5)$ | $^3\text{He}/\text{H} (10^5)$ | [ $^3\text{He}/\text{H}$ ] |
|-----------|------------------------|--------|----------------------------|-----------------------------------|-------------------------------|----------------------------|
| G1.1      | 0.0                    | -0.025 | 0.071                      | 0.45                              | 0.63                          | -0.374                     |
| G23.4     | 3.6                    | 0.124  | 0.069                      | 0.71                              | 1.03                          | -0.164                     |
| G29.9     | 4.4                    | 0.049  | 0.067                      | 2.04                              | 3.05                          | 0.307                      |
| G20.73    | 4.9                    | 0.124  | 0.058                      | 1.06                              | 1.83                          | 0.086                      |
| M16N      | 6.6                    | 0.124  | 0.082                      | 1.30                              | 1.59                          | 0.024                      |
| M16       | 6.6                    | 0.124  | 0.079                      | 1.42                              | 1.80                          | 0.079                      |
| S90       | 7.6                    | -0.100 | 0.070                      | 1.45                              | 2.07                          | 0.140                      |
| S171A     | 8.9                    | -0.174 | .....‡                     | 0.64                              | 0.64                          | -0.370                     |
| NGC7538   | 9.9                    | -0.174 | 0.084                      | 1.32                              | 1.57                          | 0.020                      |
| Rosette A | 10.0                   | -0.249 | 0.070                      | 1.83                              | 2.61                          | 0.241                      |
| Rosette B | 10.0                   | -0.249 | 0.084                      | 2.01                              | 2.39                          | 0.203                      |
| S252      | 10.0                   | -0.472 | .....‡                     | 0.98                              | 0.98                          | -0.185                     |
| G133.8    | 10.1                   | -0.174 | 0.081                      | 1.73                              | 2.14                          | 0.153                      |
| S311      | 11.0                   | -0.472 | 0.071                      | 1.82                              | 2.56                          | 0.233                      |
| S206      | 11.5                   | -0.323 | 0.092                      | 1.89                              | 2.05                          | 0.137                      |
| G93.06    | 11.8                   | -0.398 | 0.068                      | 1.15                              | 1.69                          | 0.052                      |
| S159      | 12.3                   | -0.323 | 0.090                      | 0.79                              | 0.88                          | -0.233                     |
| G201.6    | 12.6                   | -0.323 | 0.082                      | 1.80                              | 2.20                          | 0.165                      |
| S298      | 13.6                   | .....§ | .....‡                     | 2.33                              | 2.33                          | 0.191                      |
| S209      | 16.2                   | -0.547 | 0.082                      | 0.90                              | 1.10                          | -0.136                     |
| S209S     | 16.2                   | -0.323 | .....‡                     | 1.40                              | 1.40                          | -0.030                     |

The [ $^3\text{He}/\text{H}$ ] abundance ratios by number for the sample of simple H II regions. The  $^3\text{He}$  plateau abundance,  $^3\text{He}/\text{H} = (1.9 \pm 0.6) \times 10^{-5}$  (s.e.m.), is defined to be the average abundance for the 16 sources with ionization corrections. The Galactic Centre source G1.1 has also been excluded from this average because the Galactic Centre vicinity may have had a different evolutionary history from the disk. We have modelled all sources as single homogeneous spheres in local thermodynamic equilibrium<sup>10</sup>. The continuum emission is used to determine the  $\text{H}^+$  column density. The [O/H] abundance is derived from the nebular electron temperature<sup>28</sup>, which in turn is set by the H91 $\alpha$  recombination line-to-continuum ratio. The source distances are either determined by using spectrophotometry on stars in or near the H II region or estimated kinematically. If necessary, an ionization correction is applied using the  $^4\text{He}^+/\text{H}^+$  abundance ratio from the observed H91 $\alpha$  and He91 $\alpha$  recombination transitions. Specifically, if the  $^4\text{He}^+/\text{H}^+$  abundance ratio in a source falls below 0.1, the average Galactic value, we deem that the source is underionized and contains neutral helium and thus scale the  $^3\text{He}/\text{H}$  abundance ratio upward. This correction should properly account for any neutral helium so long as the fraction of helium ionization does not vary significantly on the angular scale of our three-arcminute telescope beam.

\* [O/H] =  $\log(\text{O}/\text{H}) - \log(\text{O}/\text{H})_{\odot}$  where  $(\text{O}/\text{H})_{\odot} = 6.3 \times 10^{-4}$ .

† [ $^3\text{He}/\text{H}$ ] =  $\log(^3\text{He}/\text{H}) - \log(^3\text{He}/\text{H})_{\odot}$  where  $(^3\text{He}/\text{H})_{\odot} = 1.5 \times 10^{-5}$ .

‡  $^4\text{He}^+$  recombination line too weak to allow for an ionization correction.

§ Weak radio continuum gives highly uncertain  $\text{H}^+$ .

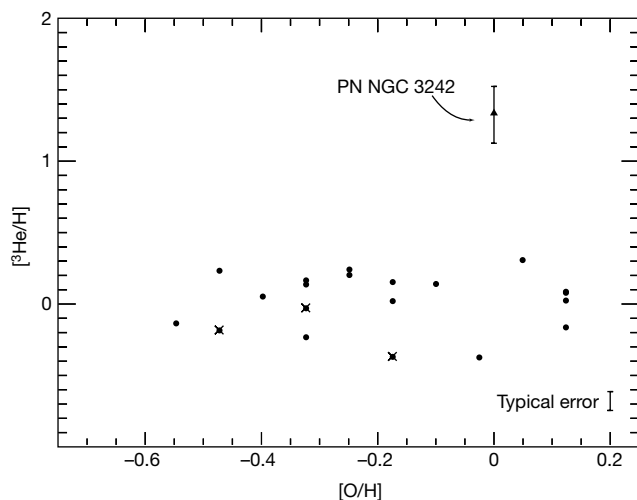


abundance, together with those derived for the other detected planetary nebulae sources, is consistent with  $^3\text{He}$  production in significant quantities by a yet to be determined fraction of stars with masses in the range  $1\text{--}2\ M_{\odot}$ , where  $M_{\odot}$  is the mass of the Sun. Specifically, the detections and upper limits on the  $^3\text{He}$  abundance for all the planetary nebulae in our sample are consistent with standard stellar production,  $^3\text{He}/\text{H} \approx 10^{-4}$ . We anticipate that the new NRAO Green Bank telescope will give us the sensitivity we need to characterize better the  $^3\text{He}$  abundances in a larger sample of planetary nebulae.

The figures show the  $[^3\text{He}/\text{H}]$  abundance ratio plotted as a function both of the source's  $[\text{O}/\text{H}]$  metallicity (Fig. 1) and of galactocentric radius,  $R_{\text{gal}}$  (Fig. 2). The simple H II regions show no trend of abundance with metallicity and no evidence for a  $^3\text{He}/\text{H}$  abundance gradient in the Galaxy. In contrast, at least some low-mass stars have produced significant amounts of  $^3\text{He}^+$ , as is shown by the planetary nebula NGC3242 abundance. Yet observations of  $^3\text{He}$  abundances in protosolar material (PSM)<sup>16</sup>, the local interstellar medium (LISM)<sup>17</sup>, and Galactic H II regions<sup>6,10</sup> are 1–2 orders of magnitude lower than those measured in planetary nebulae. Moreover, when the  $^3\text{He}$  abundances are compared with that of PSM and the LISM, there is no evidence for any stellar enrichment during the last 4.5 Gyr. This confusing situation—stars are observed to produce  $^3\text{He}$ , but then this  $^3\text{He}$  does not show up elsewhere—is called “the  $^3\text{He}$  problem” (ref. 8).

These results strongly constrain Galactic chemical evolution models. All models that adopt standard stellar  $^3\text{He}$  nucleosynthesis overproduce  $^3\text{He}$ . All realistic Galactic evolution models which match the other observational constraints predict  $^3\text{He}$  abundances that are inconsistent with those observed both locally and globally in the Milky Way unless they adopt alternative nucleosynthesis with a strongly reduced  $^3\text{He}$  contribution from low- and intermediate-mass stars<sup>15,18</sup>.

This evidence that the  $^3\text{He}$  produced during the main-sequence evolution of low-mass stars is not reaching the interstellar medium may be related to the anomalously low carbon isotopic ratios observed in low-mass red giants<sup>3,11,19,20</sup>. If extra internal mixing driven by stellar rotation occurs during the evolution of low-mass red giant stars then  $^3\text{He}$  could be depleted<sup>12</sup>. Observed  $^{12}\text{C}/^{13}\text{C}$  ratios suggest that about 90% of evolved low-mass stars go through this

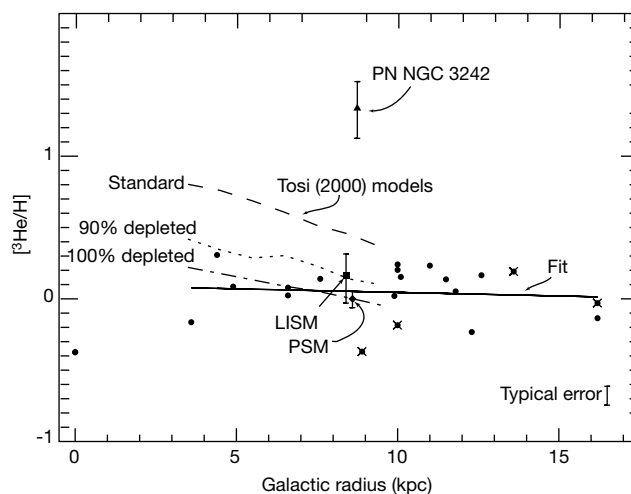


**Figure 1** Interstellar  $^3\text{He}/\text{H}$  abundances as a function of source metallicity. The  $[^3\text{He}/\text{H}]$  abundances by number derived for the “simple” H II region sample are given with respect to the solar ratio (see Table 1). The typical error for the H II region abundances (circles) is shown in the right-hand corner. Also shown is the abundance derived for the planetary nebula NGC3242 (triangle). We note that there is no trend in the  $^3\text{He}/\text{H}$  abundance with source metallicity. PN, planetary nebula.

stage and deplete at least some of their  $^3\text{He}$  (ref. 13). For perhaps 10% of the stars the rotation is below some critical value and mixing does not occur— $^{12}\text{C}/^{13}\text{C}$  remains high and  $^3\text{He}$  survives<sup>14</sup>. We note that even those stars with a low  $^{12}\text{C}/^{13}\text{C}$  and depleted  $^3\text{He}$  still have a  $^3\text{He}$  abundance which is comparable to or slightly larger than the initial value at stellar birth. In terms of chemical evolution, such stars are ‘non-producers’ rather than ‘destroyers’ of  $^3\text{He}$ .

Alternatively, Sackmann and Boothroyd<sup>21</sup> incorporate extra mixing in stellar models in a parametrized way. They find some parts of parameter space in which  $^3\text{He}$  is destroyed. These stellar models are the ones used in the chemical evolution calculations shown in Fig. 2. Even in the most extreme case, wherein all stars undergo mixing, there is still a gradient in the  $^3\text{He}/\text{H}$  abundance with respect to  $R_{\text{gal}}$ . This shows that, averaged over all masses, the stellar population is still a net producer of  $^3\text{He}$ . Moreover, any mixing scenario and chemical evolution model must ultimately account for the fact that around 10% of stars are observed not to have engaged in extra mixing and will unavoidably add some  $^3\text{He}$  to the interstellar medium.

Thus, observations of  $^3\text{He}$  (Galactic H II regions, PSM and LISM), observations of the  $^{12}\text{C}/^{13}\text{C}$  abundance ratio in red giant stars, and extra mixing stellar models all indicate that: (1) the stellar contribution to the  $^3\text{He}$  abundance evolution is positive (that is,  $^3\text{He}$  increases), and (2)  $^3\text{He}$  interstellar medium enrichment is small compared to the size of the present observational errors. We now believe that we can estimate the primordial  $^3\text{He}$  abundance because the existence of ‘the  $^3\text{He}$  plateau’ (ref. 6) for our sample of large,



**Figure 2** Interstellar  $^3\text{He}/\text{H}$  abundances as a function of Galactic radius. Plot parameters are the same as in Fig. 1. We note that there is no trend in the  $^3\text{He}/\text{H}$  abundance with Galactic position. The dispersion of abundances is very small—comparable to that of the famous “lithium plateau”<sup>29</sup>. Also shown are the abundances derived for the local interstellar medium (LISM, square), and the protosolar material (PSM, diamond). The  $^4\text{He}$  recombination lines for the four sources shown as crossed circles are too weak for an ionization correction to be made with current data. Because these corrections are generally small and because of the strategic location of these sources in the Galaxy we felt it desirable to include them. The best linear fit to the H II region data is shown as a solid line. The G1.1 nebula is not included in the fit because the Galactic Centre region is thought to have a different chemical enrichment history from the disk. The predictions of several models for Galactic chemical evolution are shown<sup>18</sup>. Together these models span the range of possibilities for  $^3\text{He}$  enrichment by stellar nuclear processing: (1) standard stellar yields (dashed); (2) 90% of all stars do not produce  $^3\text{He}$  (dotted); and (3) 100% of all stars do not produce  $^3\text{He}$  (dash-dotted). Altogether, these H II region abundances show no evidence for substantial  $^3\text{He}$  interstellar medium enrichment by any process.

diffuse Galactic H II regions suggests that there is a floor value to the  $^3\text{He}/\text{H}$  abundance. If, as seems plausible, stellar processing enriches the interstellar medium's  $^3\text{He}$  abundance by a positive (but perhaps nearly zero) amount, this floor value provides an upper limit to the primordial  $^3\text{He}$  abundance produced by Big Bang nucleosynthesis (BBNS). A very conservative estimate for the floor abundance can be made by adopting the plateau value itself,  $^3\text{He}/\text{H} = (1.9 \pm 0.6) \times 10^{-5}$  (s.e.m.) by number, which can serve as a strong upper limit to the BBNS production.

Some of the scatter in the Fig. 2 abundances could be real. Some stars do produce  $^3\text{He}$ , so the upward scatter in the abundances could be partly due to stellar production. Furthermore, there is an indication that the helium is underionized in the two low-lying points at  $R_{\text{gal}} = 9$  and 10 kpc, but our  $^4\text{He}$  recombination lines are not determined well enough to allow for an ionization correction. This all suggests that our definition of the  $^3\text{He}$  plateau abundance value is a very conservative limit to the primordial value.

In principle, we could set an upper limit to the primordial  $^3\text{He}$  abundance from one well observed simple source at large  $R_{\text{gal}}$  and with high helium ionization. The H II region S209 is such a source. We observed S209 during many epochs spread over 15 years; our total integration time is 133 h. The  $^3\text{He}^+$  emission can be detected in any subset of the data that has adequate integration time and its intensity varies within statistically expected limits. We have very-high-quality continuum and recombination line data. We used this information to construct a detailed, high-quality model<sup>10</sup> showing that S209 is simple. The degree of He ionization is high and our  $^4\text{He}^+$  data are good enough to allow us to make an ionization correction. Optical observations of  $^4\text{He}^+$ , although complicated by high reddening in the direction of S209, also suggest that ionization is a minor problem<sup>22</sup>. The nearby very-low-density H II region, S209S, has an abundance that is the same as S209, within the errors. Our estimate for the S209  $^3\text{He}$  abundance is  $^3\text{He}/\text{H} = (1.1 \pm 0.2) \times 10^{-5}$ . There is always danger in basing conclusions on one object, but any average of outer Galaxy sources would mean bringing in objects whose abundances are compromised in some way compared to S209. At this point after almost 20 years of work our best upper limit on the primordial  $^3\text{He}$  abundance is  $(1.1 \pm 0.2) \times 10^{-5}$ .

Primordial abundances can be directly related to the baryon-to-photon ratio of the universe,  $\eta$ , for example, by using the fitting formulae of ref. 23. However, the expansion of the Big Bang is more commonly described in terms of a parameter  $\Omega$ , which is a fractional measure of the sources of the gravity that controls the expansion of the universe. We note that  $\Omega$  can include contributions from matter or energy. An open universe which will expand forever has  $\Omega < 1$ ; a closed universe whose expansion will eventually reverse has  $\Omega > 1$ . The amount of ordinary baryonic matter in the universe,  $\Omega_{\text{B}}$ , can be derived from  $\eta$  and the current rate of expansion as measured by the Hubble constant,  $h$ , which is measured in units of  $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ .

Our values for  $\eta$  and  $\Omega h^2$  are given in Table 2, together with some other recent determinations. Because of deuterium's simple post-Big-Bang evolution, measurements of deuterium in high-redshift quasars<sup>23</sup> may potentially give the 'cleanest' primordial abundance. However, because the largest errors are systematic and hence poorly known, a truly robust cosmological conclusion can be drawn only when concordance between different determinations is achieved. The problem with systematic errors is illustrated by the debate over estimates of the cosmological  $^4\text{He}$  abundance in the last few years. To get  $\eta$  from  $^4\text{He}$  requires very high precision. Results split into two camps, one with low  $^4\text{He}$  (ref. 24) and another with high  $^4\text{He}$  (ref. 25). Such a dichotomy clearly indicates a systematic error much larger than the internal error.

A consistent picture is now emerging from experiments attempting to constrain the parameters of Big Bang cosmology. Table 2

**Table 2 Determinations of the cosmological baryon to photon ratio**

| Method                                                          | Abundance                    | $\eta$                                | $\Omega_{\text{B}} h^2$     |
|-----------------------------------------------------------------|------------------------------|---------------------------------------|-----------------------------|
| Primordial $^3\text{He}/\text{H}^*$<br>(based on S209)          | $1.1 \pm 0.2 \times 10^{-5}$ | $5.4^{+2.2}_{-1.2} \times 10^{-10}$   | $0.020^{+0.007}_{-0.003}$   |
| $^3\text{He}/\text{H}$ plateau*                                 | $1.9 \pm 0.6 \times 10^{-5}$ | $> 2.4^{+1.8}_{-0.7} \times 10^{-10}$ | $> 0.009^{+0.014}_{-0.008}$ |
| Quasar deuterium <sup>23</sup>                                  | $3.0 \pm 0.2 \times 10^{-5}$ | $5.6 \pm 0.5 \times 10^{-10}$         | $0.020 \pm 0.002$           |
| High $^4\text{He}$ (ref. 25)<br>( $Y_{\text{p}}$ mass fraction) | $0.244 \pm 0.02$             | $4.0^{+0.9}_{-0.4} \times 10^{-10}$   | $0.015^{+0.003}_{-0.002}$   |
| Low $^4\text{He}$ (ref. 24)                                     | $0.234 \pm 0.02$             | $1.8^{+0.3}_{-0.2} \times 10^{-10}$   | $0.007 \pm 0.001$           |
| DASI CMB experiment <sup>26</sup>                               | ...                          | ...                                   | $0.022^{+0.004}_{-0.003}$   |

DASI, Degree Angular Scale Interferometer; CMB, cosmic microwave background;  $Y_{\text{p}}$ ,  $^4\text{He}/\text{H}$  mass fraction.

\* This work.

shows that the values of  $\eta$  derived from our S209 result and that from quasar deuterium are in excellent agreement and clearly fall on the side of the higher  $^4\text{He}$  abundance. Measurements of fluctuations in the cosmic microwave background radiation can also give estimates for  $\Omega_{\text{B}} h^2$ . Recent results from the Degree Angular Scale Interferometer (DASI)<sup>26</sup> experiment are also given in Table 2. These different estimates for the primordial baryon density expressed as a fraction of the closure density are in excellent agreement and give for  $h = 0.72 \pm 0.08$  (ref. 27) a value  $\Omega_{\text{B}} = 0.04$ . □

Received 24 July; accepted 26 November 2001.

- Walker, T. P., Steigman, G., Schramm, D. N., Olive, K. A. & Kang, H. S. Primordial nucleosynthesis redux. *Astrophys. J.* **376**, 51–69 (1991).
- Rood, R. T., Steigman, G. & Tinsley, B. M. Stellar production as a source of  $^3\text{He}$  in the interstellar medium. *Astrophys. J.* **207**, L57–L60 (1976).
- Rood, R. T., Bania, T. M. & Wilson, T. L. The 8.7 GHz hyperfine line of  $^3\text{He}$  in galactic H II regions. *Astrophys. J.* **280**, 629–647 (1984).
- Bania, T. M., Balser, D. S., Rood, R. T., Wilson, T. L. & Wilson, T. A.  $^3\text{He}$  in the Milky Way interstellar medium: summary of relevant observations. *Astrophys. J. (Suppl.)* **113**, 353–366 (1997).
- Balser, D. S., Bania, T. M., Rood, R. T. & Wilson, T. L.  $^3\text{He}$  abundance in planetary nebulae. *Astrophys. J.* **483**, 320–334 (1997).
- Rood, R. T., Bania, T. M., Balser, D. S. & Wilson, T. L. Helium-3: Status and prospects. *Space Sci. Rev.* **84**, 185–198 (1998).
- Rood, R. T., Bania, T. M. & Wilson, T. L. Detection of helium-3 in a planetary nebula. *Nature* **355**, 618–619 (1992).
- Galli, D., Stanghellini, L., Tosi, M. & Palla, F.  $^3\text{He}$  in planetary nebulae: A challenge to stellar evolution models. *Astrophys. J.* **477**, 210–225 (1997).
- Balser, D. S., Rood, R. T. & Bania, T. M. The  $^3\text{He}$  abundance in the planetary nebula NGC3242. *Astrophys. J.* **522**, L73–L76 (1999).
- Balser, D. S., Bania, T. M., Rood, R. T. & Wilson, T. L.  $^3\text{He}$  in the Milky Way interstellar medium: Abundance determinations. *Astrophys. J.* **510**, 759–783 (1999).
- Charbonnel, C. A coincident explanation for  $^{12}\text{C}/^{13}\text{C}$ ,  $^7\text{Li}$ , and  $^3\text{He}$  anomalies in red giant stars. *Astrophys. J.* **453**, L41–L44 (1995).
- Charbonnel, C., Brown, J. A. & Wallerstein, G. Mixing processes during the evolution of red giants with moderate metal deficiencies: the role of molecular-weight barriers. *Astron. Astrophys.* **332**, 204–214 (1998).
- Charbonnel, C. & do Nascimento, J. D. Jr. How many low-mass stars do destroy  $^3\text{He}$ ? *Astron. Astrophys.* **336**, 915–919 (1998).
- Charbonnel, C. Mixing in stars and the evolution of the  $^3\text{He}$  abundance. *Space Sci. Rev.* **84**, 199–206 (1998).
- Dearborn, D. S. P., Steigman, G. & Tosi, M. Galactic evolution of D and  $^3\text{He}$  including stellar production of  $^3\text{He}$ . *Astrophys. J.* **465**, 887–897 (1996).
- Mahaffy, P. R., Donahue, T. M., Atreya, S. K., Owen, T. C. & Niemann, H. B. Galileo probe measurements of D/H and  $^3\text{He}/\text{H}$  in Jupiter's atmosphere. *Space Sci. Rev.* **84**, 251–263 (1998).
- Gloeckler, G. & Geiss, J. Measurement of the abundance of helium-3 in the Sun and local interstellar cloud with SWICS on *Ulysses*. *Space Sci. Rev.* **84**, 275–284 (1998).
- Tosi, M. in *The Light Elements and Their Evolution* (eds da Silva, L., Spite, M. & de Medeiros, J. R.) 525–534 (Astronomical Society of the Pacific, ASP, San Francisco, 2000).
- Wasserburg, G. J., Boothroyd, A. I. & Sackmann, I.-J. Deep circulation in red giant stars: A solution to the carbon and oxygen isotope puzzles? *Astrophys. J.* **447**, L37–L40 (1995).
- Hogan, C. J. Giant branch mixing and the ultimate fate of primordial deuterium in the Galaxy. *Astrophys. J.* **441**, L17–L20 (1995).
- Sackmann, I.-J. & Boothroyd, A. I. Creation of  $^7\text{Li}$  and destruction of  $^3\text{He}$ ,  $^9\text{Be}$ ,  $^{10}\text{B}$ , and  $^{11}\text{B}$  in low-mass red giants due to deep circulation. *Astrophys. J.* **510**, 217–231 (1999).
- Deharveng, L., Peña, M., Caplan, J. & Costero, R. Oxygen and helium abundances in Galactic H II regions—II. Abundance gradients. *Mon. Not. R. Astron. Soc.* **311**, 329–345 (2000).
- Burles, S., Nollett, K. M. & Turner, M. S. Big-bang nucleosynthesis predictions for precision cosmology. *Astrophys. J.* **552**, L1–L5 (2001).
- Olive, K. A., Steigman, G. & Skillman, E. D. The primordial abundance of  $^4\text{He}$ : an update. *Astrophys. J.* **483**, 788–797 (1997).
- Izotov, Y. I. & Thuan, T. X. The primordial abundance of  $^4\text{He}$  revisited. *Astrophys. J.* **500**, 188–216 (1998).
- Pryke, C. et al. Cosmological parameter extraction from the first season of observations with DASI. *Astrophys. J.* (submitted); reprint astro-ph/0104490 at (<http://xxx.lanl.gov>) (2001).

27. Freedman, W. L. *et al.* Final results from the Hubble space telescope key project to measure the Hubble constant. *Astrophys. J.* **553**, 47–72 (2001).
28. Shaver, P. A., McGee, R. X., Newton, L. M., Danks, A. C. & Pottasch, S. R. The galactic abundance gradient. *Mon. Not. R. Astron. Soc.* **204**, 53–112 (1983).
29. Spite, M. & Spite, F. Lithium abundance at the formation of the galaxy. *Nature* **297**, 483–485 (1982).

# Acknowledgements

We thank the international light elements community for support over the years.

Correspondence and requests for materials should be addressed to T.M.B. (e-mail: bania@bu.edu).

# Interstellar scintillation as the origin of the rapid radio variability of the quasar J1819+3845

J. Dennett-Thorpe\*† & A. G. de Bruyn\*‡

\* Kapteyn Astronomical Institute, University of Groningen, 9700 AV, Groningen, The Netherlands

† Astronomical Institute Anton Pannekoek, University of Amsterdam, 1098 SJ, Amsterdam, The Netherlands

‡ ASTRON, PO Box 2, 7990 AA, Dwingeloo, The Netherlands

The liberation of gravitational energy as matter falls onto a supermassive black hole at the centre of a galaxy is believed to explain the high luminosity of quasars. The variability of this emission from quasars and other types of active galactic nuclei can provide information on the size of the emitting regions and the physical process of fuelling the black hole. Some active galactic nuclei are variable at optical (and shorter) wavelengths, and display radio outbursts over years and decades. These active galactic nuclei often also show faster intraday variability at radio wavelengths<sup>3,4</sup>. The origin of this rapid variability has been extensively debated<sup>5</sup>, but a correlation between optical and radio variations in some sources<sup>6,7</sup> suggests that both are intrinsic. This would, however, require radiation brightness temperatures that seem physically implausible, leading to the suggestion that the rapid variations are caused by scattering of the emission by the interstellar medium inside our Galaxy<sup>8,9</sup>. Here we show that the rapid variations in the extreme case of quasar J1819+3845 (ref. 10) indeed arise from interstellar scintillation. The transverse velocity of the scattering material reveals the presence of plasma with a surprisingly high velocity close to the Solar System.

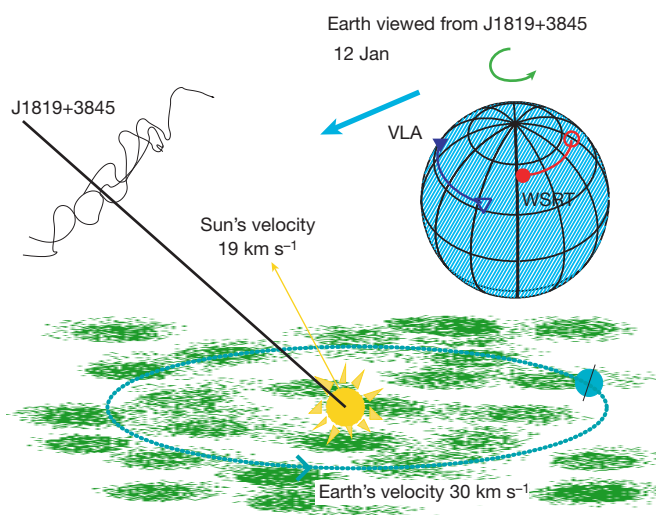
The most variable extragalactic source known at radio wavelengths is the quasar J1819+3845. This quasar shows variations of factors of 4 or more on a timescale of hours, easiest interpreted as scintillation due to scattering in the interstellar medium (ISM)<sup>10</sup>. One can think of this scattering medium as focusing and defocusing the waves from the source, producing a pattern of dark and bright patches<sup>11</sup>. As we move through these patches we observe a temporal variation in intensity. In order to display this effect, the source must be small; accordingly J1819+3845 is only tens of microarcseconds in angular size (a few light months at the source redshift 0.54)<sup>10</sup>.

As the Earth moves through the scintillation pattern, two telescopes at different locations should see a difference in the observed time of the intensity maxima and minima (Fig. 1). A similar technique was used both by Rickett and Lang<sup>12</sup>—who used the decorrelation of signals between telescopes to put limits on the coherence size and apparent velocity of the scintillation patches

from a pulsar—and by Jauncey *et al.*<sup>13</sup> for the other known intra-hour variable PKS0405-387<sup>14</sup>. Compared to other intra-day variable (IDV) sources, or even PKS0405-387, the rapidity and depth of the variations in J1819+3845 make it the perfect candidate for measurement of this delay.

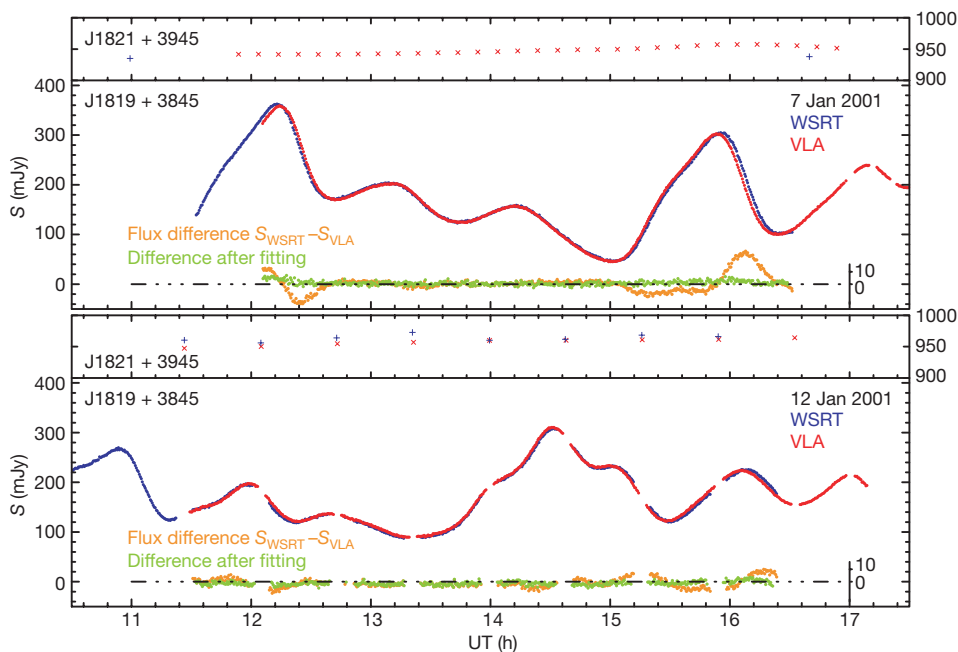
The observations presented here were conducted simultaneously with the National Radio Astronomy Observatory (NRAO) Very Large Array (VLA) in New Mexico, and the Westerbork Synthesis Radio Telescope (WSRT) in the Netherlands. The source showed large, rapid variations on both days (Fig. 2). We observed a difference in the time at which the earliest peaks reach the telescopes (WSRT leads VLA). We also noted that at the end of the observations this delay reversed, such that the VLA leads WSRT. This delay, and reversal, occurred on both days. If the scattering plasma is stationary with respect to the local standard of rest (LSR), the predicted time lag from the motion of the Earth and Sun alone is  $\sim 200$  s, with the signal arriving at the VLA first throughout the run. This clearly cannot account for the observations, so we used the observations to solve for the velocity of the scattering plasma with respect to the LSR, transverse to the line of sight.

From a minimization of the total squared flux density difference we calculate the projected plasma velocity. We obtain for 7 Jan 2001:  $v_{\text{RA}} = -32.8 \pm 0.5 \text{ km s}^{-1}$ ,  $v_{\text{dec}} = 14.3 \pm 3 \text{ km s}^{-1}$  and for 12 Jan 2001:  $v_{\text{RA}} = -32.2 \pm 0.5 \text{ km s}^{-1}$ ,  $v_{\text{dec}} = 16.1 \pm 1 \text{ km s}^{-1}$ . These values show that we find very good agreement between the results obtained from observations of both days. (This implies that the assumption of a temporally ‘frozen’ scintillation pattern is justified.) Combining these, we calculate the plasma velocity to be  $v_{\text{RA}} = -32.5 \pm 0.5 \text{ km s}^{-1}$ ,  $v_{\text{dec}} = 15.5 \pm 1 \text{ km s}^{-1}$  (that is, the plasma moves towards west and north). We also fitted a single time delay to  $\sim 15$  minutes of data around the largest maxima, normalizing the two light curves over the time of interest to the same mean flux density. Solutions which pass through these values are consistent with the results quoted above. We plot both these results and the delays fitted as described above in Fig. 3. We can constrain the plasma velocity so well because at these values the predicted delay



**Figure 1** Diagram showing the projection of the interstellar scintillation onto the Solar System. The inset shows the Earth and its projected velocity with respect to the LSR (blue arrow) as seen from J1819+3845. Filled and open symbols show the telescopes at the start and the end of the observation run, respectively. As the Earth rotates, the value of the ‘delay’ in arrival of intensity extrema will change during the observation due to the changing orientation of the baseline between telescopes with respect to the vector velocity of the scintillation pattern. The delay also depends on the time of year (due to the Earth’s movement around the Sun) and the relative motion of the Solar System and scattering plasma.





**Figure 2** Results of VLA and WSRT observations on 7 and 12 January 2001. Data were averaged over 30 s. The time difference in the arrival of the signals can be seen as a horizontal separation of the blue and red lines. The vertical axis of the difference signal before and after fitting has been expanded by a factor of 2.5. A total of 11 VLA antennas were used for the observations described here. The central frequency at both telescopes was 4,875 MHz, and the bandwidths similar (80 and 100 MHz). VLA observations of a nearby calibrator source (J1821+3945) are shown in the top panel of each plot. Slightly different observation strategies were used on each run. On 7 January 2001 we observed the source continuously, with a sub-set of antennas observing the calibrator. On 12 January 2001 we used the same antennas to nod between the source and the calibrator, observing the latter for 3 minutes every  $\sim 35$  minutes. On both dates, the observations

described above were bracketed with observations of the calibrator and a flux density calibrator (3C286). Elevation and other time-dependent gain variations at the VLA determined using the calibrator J1821+3945 were up to 3%. The data were corrected for these effects (see below). Observations in good weather at WSRT typically yield 1% flux density reliability. This is applicable to all but the first hour of the 7 January 2001 run, when light rain caused an estimated extra 1–2% extinction. The error we quote on the result is obtained using a number of different determinations of the absolute flux density scale, and acceptable changes throughout the run, including: no temporal changes in gain corrections, a linear change, and gain corrections determined by observations of the calibrator, which is assumed stable. Thermal noise in 30 s is 1 mJy (WSRT) and 0.6 mJy (VLA).

changes very rapidly with changes in plasma velocity. The existence of many maxima and minima in our data is important as it removes the possibility of the delay—and the calculated plasma velocity—being an artefact of flux calibration errors.

We have previously suggested<sup>15</sup> a velocity of the plasma screen in order to explain the change in timescale of the variations over a year. However, this was not a measurement but a mere interpretation of the data, and the velocity was poorly constrained. The preliminary result presented in ref. 15,  $v_{\text{RA}} = 3 \text{ km s}^{-1}$ ,  $v_{\text{dec}} = 25 \text{ km s}^{-1}$ , predicts a very different delay: from around +60 s at the beginning of the observations to around +150 s at the end.

We point out that we are measuring the transverse velocity of the local plasma, whereas in general, radial velocities are measured (using spectroscopic techniques). Also note that whereas scintillation observations of pulsars are significantly affected by their own transverse velocities (up to several hundreds of  $\text{km s}^{-1}$ ), the effect of a transverse velocity, even as high as the speed of light, of jet features in active galactic nuclei (AGNs) is reduced to a fraction of a  $\text{km s}^{-1}$  by the distances involved.

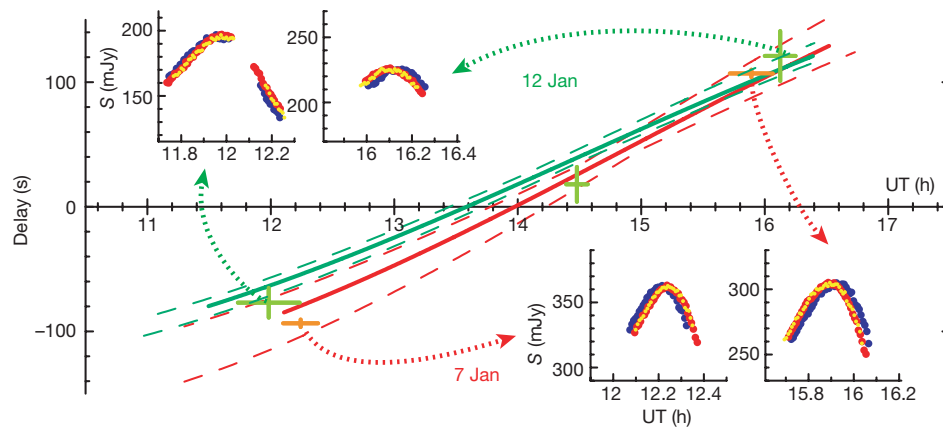
The determination of a peculiar velocity of the plasma argues that the scattering structure is discrete. This measurement of the transverse velocity of the ionized scattering material cannot be directly compared to radial velocity measurements of H<sub>I</sub>, but we note that there is no anomalous feature known in this direction<sup>16</sup>. A nearby pulsar J1813+4013 has unremarkable dispersion and rotation measures ( $42 \text{ cm}^{-3} \text{ pc}$  and  $47 \text{ rad m}^{-2}$ )<sup>17</sup>, although the medium in front of J1819+3845 may not extend to this line of sight. We have previously estimated<sup>10</sup> that this scattering region is located at around 20 pc from the Sun. The scatterer may be related to the Local Bubble<sup>18</sup>, and if so, for the reasons given above, is more likely

to be a discrete structure at the boundary than the putative intervening hot gas.

A small difference in the flux densities observed at the two telescopes (after delay correction) as the telescopes pass through different regions of scintillation is expected. The size of the scintillation patches in the direction of motion is  $\sim 4,000 \text{ s} \times 25 \text{ km s}^{-1} = 10^5 \text{ km}$ . If the patches were of similar size in the direction perpendicular to this, we can estimate the difference in the intensities on a typical baseline of 6,000 km by considering the flux density differences between the two curves if they are shifted  $6,000/10^5 \times 4,000 = 240 \text{ s}$ . This gives a potential 5–10% flux density difference (after delay correction) between the two curves. However, differences are less than 2% on both days (Fig. 2), suggesting that the scintillation pattern is highly anisotropic, in agreement with preliminary analysis of the variation of the timescale over a year<sup>15</sup>.

These observations may be explained by an unusually high fraction of turbulent electrons, if the electron density is that of the typical extended ISM. However, in comparison to the conditions found in highly turbulent, ionized regions, for example around OB stars<sup>19–21</sup>, the required conditions are very modest. The A0 star  $\alpha$  Lyrae (Vega), which at 8 pc is within the estimated distance range for the screen, is at a projected angular separation from J1819+3845 of  $3^\circ$ , which is only 0.4 pc. Whether the origin of the turbulent plasma could be related to mass loss from this star, or is a feature of the so-called Local Fluff<sup>22</sup>, remains a subject for further investigation.

Our observations prove that the variations in J1819+3845, the most extreme example of an IDV source, are due to scintillation. Yet, as this source must be extremely small, it could easily be intrinsically variable over timescales of a few months (using arguments based on



**Figure 3** The calculated delays as a function of observing time for both days of observation. The delays are calculated as a function of time from the projected baseline and velocity vector. We use a circular Earth orbit, and assume that the Sun is moving at  $19.7 \text{ km s}^{-1}$  towards right ascension  $18 \text{ h}$ , declination  $30^\circ$  (J2000). As the direction of the Sun's motion happens to be only  $10^\circ$  from J1819+3845, errors caused by its uncertainty are negligible. For a wide range of plasma velocities we calculated the total squared flux density difference, after time shifting the data according to the calculated delay. Minimizing this quantity gives the plasma velocity. The associated delays are plotted as

diagonal lines. The results from a given calibration scheme produce errors of less than  $1 \text{ km s}^{-1}$ , but the final error is dominated by calibration uncertainty. The error we quote is that obtained using a number of different determinations of the absolute flux density scale, and acceptable changes throughout the run. In addition, independent fits to portions of the data are shown with  $3\sigma$  error bars. The data from times with large delays are shown in the insets (red and dark blue as before, yellow showing the WSRT data after fitting).

the travel time of light). Consideration of relativistic outflow could reduce the timescale to weeks. Intrinsic variations on these timescales would be difficult to separate from scintillation-induced variations. We have evidence, however, that the source is remarkably stable over a period of many months, with only a slow increase in brightness over the past two years (J.D.-T. and A.G.d.B., manuscript in preparation). Scintillation in known variable sources and those with slower variations will be more difficult to prove directly. However, the few percent variations seen on day scales in many IDV sources<sup>23</sup> are likely to result from scattering of structures of tens of microarcseconds through more typical ISM<sup>24</sup>. This implies that most AGN have brightness temperatures no higher than  $10^{12} - 10^{14} \text{ K}$ , not  $10^{17} - 10^{20} \text{ K}$ , as would otherwise be required. Perhaps the most serious flaw in an intrinsic explanation of the IDV phenomenon, which requires extremely small angular sizes, is that the scintillation-induced variations are predicted to dominate all the observed variability.

The lack of correlation between the strength or timescale of variations and the frequency or galactic latitude is sometimes cited as evidence against interstellar scintillation as cause for IDV (ref. 25). However, as the variations are affected by the intrinsic size of the source, the location and motion of the effective scatterer and the season of the observations, such correlations may easily be obscured. Although the annual modulation in the variations of B0917+62 have been explained in terms of scintillation through the general Galactic ISM<sup>26,27</sup>, our results on J1819+2845 show that one cannot assume that the scattering effectively occurs throughout this medium. It is noteworthy that the scattering of resolved AGN will be weighted by the material closer to the observer.

The delay in arrival times of the intensity variations at two widely separated radio telescopes proves that a propagation effect causes the extreme variations in the J1819+3845 quasar. The small size of this quasar is probably related to the unusual rising radio spectrum. The small number of sources scintillating in this manner is presumably due to a combination of an unusually small, stable source, and an atypically strong, close scatterer. However, we expect that other similar extreme IDVs will be discovered. □

Received 12 July; accepted 2 November 2001.

1. Rees, M. J. Black hole models for active galactic nuclei. *Annu. Rev. Astron. Astrophys.* **22**, 471–506 (1984).

2. Aller, H. D., Aller, M. E., Latimer, G. E. & Hodge, P. E. Spectra and linear polarizations of extragalactic variable sources at centimeter wavelengths. *Astrophys. J. Suppl. Ser.* **59**, 513–768 (1985).
3. Heeschen, D. S. Flickering of extragalactic radio sources. *Astron. J.* **89**, 1111–1123 (1984).
4. Quirrenbach, A., Witzel, A., Krichbaum, T., Hummel, C. A. & Alberdi, A. Rapid variability of extragalactic radio sources. *Nature* **337**, 442–444 (1989).
5. Wagner, S. J. & Witzel, A. Intraday variability in quasars and BL Lac objects. *Annu. Rev. Astron. Astrophys.* **33**, 163–198 (1995).
6. Wagner, S. J. *et al.* Rapid variability in S5 0716+714 across the electromagnetic spectrum. *Astron. J.* **111**, 2187–2211 (1996).
7. Peng, B. *et al.* Infrared, radio and optical variability of the BL Lacertae object 2007+777. *Astron. Astrophys.* **353**, 937–943 (2000).
8. Rickett, B. J., Coles, W. A. & Bourgois, G. Slow scintillation in the interstellar medium. *Astron. Astrophys.* **134**, 390–395 (1984).
9. Blandford, R., Narayan, R. & Romani, R. W. Flicker of extragalactic radio sources and refractive interstellar scintillation. *Astrophys. J.* **301**, L53–L56 (1986).
10. Dennett-Thorpe, J. & de Bruyn, A. G. The discovery of a microarcsecond quasar: J1819+3845. *Astrophys. J.* **529**, L65–L68 (2000).
11. Narayan, R. The physics of pulsar scintillation. *Phil. Trans. R. Soc. Lond. A* **341**, 151–165 (1992).
12. Lang, K. R. & Rickett, B. J. Size and motion of the interstellar scintillation pattern from observations of CP1133. *Nature* **225**, 528–530 (1970).
13. Jauncey, D. L. *et al.* in *Astrophysical Phenomena Revealed by Space VLBI, Proceedings of the VSOP Symposium, 2000* (eds Hirabayashi, H., Edwards, P. G. & Murphy, D. W.) 147–150 (Institute of Space and Astronautical Science, Tokyo, 2000).
14. Kedziora-Chudczer, L. *et al.* PKS 0405–385: The smallest radio quasar? *Astrophys. J.* **490**, L9–L12 (1997).
15. Dennett-Thorpe, J. & de Bruyn, A. G. in *Galaxies and their Constituents at the Highest Angular Resolution* (eds Schilizzi, R. T. *et al.*) 88–89 (IAU Symp. 205, Astronomical Society of the Pacific, San Francisco, 2000).
16. Hartmann, D. & Burton, W. *Atlas of Galactic Neutral Hydrogen* (Cambridge University Press, Cambridge, 1997).
17. Hamilton, P. A. & Lyne, A. G. Faraday rotation measurements on 163 pulsars. *Mon. Not. R. Astron. Soc.* **224**, 1073–1081 (1987).
18. Bhat, N. D. R., Gupta, Y. & Rao, A. P. Pulsar scintillation and the Local Bubble. *Astrophys. J.* **500**, 262–279 (1998).
19. Taylor, J. H. & Cordes, J. M. Pulsar distances and the galactic distribution of free electrons. *Astrophys. J.* **411**, 674–684 (1993).
20. Cordes, J. M., Weisberg, J. M. & Boriakoff, V. Small-scale electron density turbulence in the interstellar medium. *Astrophys. J.* **288**, 221–247 (1985).
21. Moran, J. M., Greene, B., Rodriguez, L. F. & Backer, D. C. The large scattering disk of NGC 6334B. *Astrophys. J.* **348**, L47–L52 (1990).
22. Cox, D. P. & Reynolds, R. J. The local interstellar medium. *Annu. Rev. Astron. Astrophys.* **25**, 303–344 (1987).
23. Quirrenbach, A. *et al.* Statistics of intraday variability in extragalactic radio sources. *Astron. Astrophys.* **258**, 279–284 (1992).
24. Rickett, B. J., Quirrenbach, A., Wegner, R., Krichbaum, T. P. & Witzel, A. Interstellar scintillation of the radio source 0917+624. *Astron. Astrophys.* **293**, 479–492 (1995).
25. Kraus, A., Witzel, A. & Krichbaum, T. P. Intraday radio variability in active galactic nuclei. *New Astron. Rev.* **43**, 685–689 (1999).
26. Rickett, B. J., Witzel, A., Kraus, A., Krichbaum, T. P. & Qian, S. J. Annual modulation in the intraday variability of quasar 0917+624 due to interstellar scintillation. *Astrophys. J.* **550**, L11–L14 (2001).
27. Jauncey, D. L. & Macquart, J.-P. Intra-day variability and the interstellar medium towards 0917+624. *Astron. Astrophys.* **370**, 9–12 (2001).

## Acknowledgements

We thank F. Briggs for the code used to predict the delays. This work was funded by the EU TMR network 'CERES'. The NRAO is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities. The WSRT operated by ASTRON is supported by the Netherlands Organization for Scientific Research (NWO).

## Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.G.B. (e-mail: ger@astron.nl).

# Mesoscopic superconductor as a ballistic quantum switch

A. S. Mel'nikov\*† & V. M. Vinokur†

\* Institute for Physics of Microstructures, Russian Academy of Sciences, 603950, Nizhny Novgorod, GSP-105, Russia

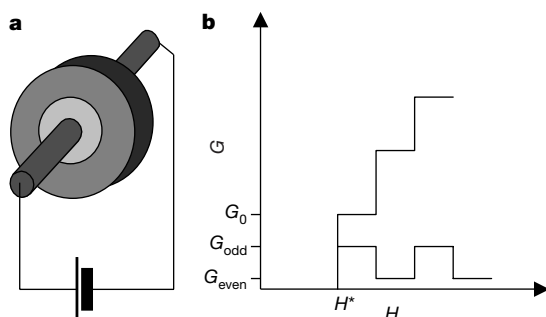
† Argonne National Laboratory, Argonne, Illinois 60439, USA

Several key experiments<sup>1–3</sup> have revealed a rich variety of vortex structures in mesoscopic superconductors in which only a few quanta of magnetic flux are trapped: these structures are polygon-like vortex 'molecules' and multi-quanta giant vortices. Ginzburg–Landau calculations<sup>4</sup> confirmed second-order phase transitions between the giant vortex states and stable molecule-like configurations<sup>5</sup>. Here we study theoretically the electronic structure and the related phase-coherent transport properties of such mesoscopic superconductor systems. The quasiparticle excitations in the vortices form coherent quantum-mechanical states that offer the possibility of controlling the phase-coherent transport through the sample by changing the number of trapped flux quanta and their configuration. The sample conductance measured in the direction of the applied magnetic field is determined by the transparency of multi-vortex configurations, which form a set of quantum channels. The transmission coefficient for each channel is controlled by multiple Andreev reflections within the

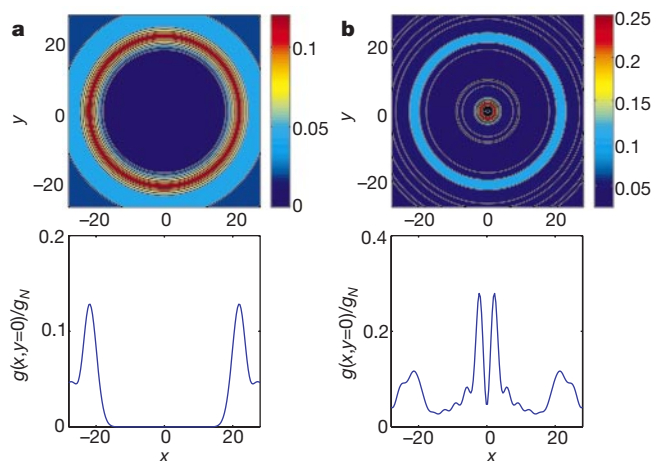
vortex cores and at the sample edge. These interference phenomena result in a stepwise behaviour of the conductance as a function of the applied magnetic field, and we propose to exploit this effect to realize a vortex-based quantum switch where the magnetic field plays the role of the gate voltage.

Giaever was the first to notice<sup>6</sup> that when magnetic flux gets trapped in a superconductor, the small normal areas that appear in parallel with superconducting areas influence transport characteristics. We find that phase-coherent transport mediated by the quasiparticle Andreev states associated with these normal domains enables us to tune the probability of tunnelling through a superconductor by changing the external magnetic field. We show that owing to a peculiar parity effect in the spatial distribution of the quasiparticle density of states (DOS), the ballistic conductance of a thin superconducting disk squeezed between very sharp contacts placed at the very centre of the disk (see Fig. 1) can alternate between finite and near-zero values as function of magnetic field. (The quasiparticle DOS has the form of a set of coaxial cylinders of finite radii, with an additional central spot if the number of trapped flux quanta is odd; the centre is hollow when the number is even.) In this regime, the mesoscopic superconductor realizes a quantum vortex switch where the external magnetic field plays the role of gate voltage.

We consider a thin disk of thickness  $d < \lambda$  (where  $\lambda$  is the London penetration depth) and radius  $R \ll \lambda_{\text{eff}} = \lambda^2/d$ . The structure of electronic states is determined by the interplay between the quasiparticle excitations bound within the cores of trapped vortices, which can either merge into multi-quanta vortices or form vortex molecules, and quasiparticle states bound near the disk edge. We start with the quasiparticle states of a giant vortex with a certain winding number  $m$ . In macroscopic samples, multi-quanta vortices are energetically unfavourable owing to the strong repulsion between the singly quantized vortices. But in mesoscopic samples, multi-quanta formations can be stable even at  $H \ll H_{c2}$ , because Meissner currents push vortices present in the sample close together. Here  $H_{c2}$  is the upper critical field. The winding number  $m = 1$  corresponds to a conventional single vortex which carries a single flux quantum  $\Phi_0 = \pi\hbar c/|e|$ . Low-lying core quasiparticle states have been found<sup>7</sup>, and can be viewed as the formation of standing quasiparticle waves owing to Andreev reflection of quasiparticles from the superconducting gap profile  $\Delta(\mathbf{r})$  confining the vortex core: the quantitative theory of the quasiparticle states is based on the Bogolubov–de Gennes equations. In  $s$ -superconductors, the



**Figure 1** Phase-coherent transport through a few-flux-quanta superconductor. **a**, Mesoscopic superconducting disk placed between the two normal metal contacts; **b**, ballistic conductance of a mesoscopic disk versus magnetic field for large-area contacts (step-like growth of conductance) and point contacts with area less than the core radius (alternating behaviour of conductance).  $G_0$  is a single vortex ballistic conductance for a large-area contact,  $G_{\text{even}}$  ( $G_{\text{odd}}$ ) is a ballistic conductance for giant vortices with an even (odd) winding number measured by a small-area contacts positioned at the disk centre.  $H^*$  is the field of the first vortex entry.



**Figure 2** Spatial distribution of the thermodynamically averaged local DOS  $g(\mathbf{r})$  normalized by its normal state value  $g_N$  for multi-quanta vortices. **a**,  $m = 2$ ; **b**,  $m = 3$ . The  $x, y$  coordinates are measured in  $1/k_F$ ,  $T \approx \Delta/k_F\xi$ ; for  $m = 2$ ,  $k_F\xi_c = 20\sqrt{2}$ ; for  $m = 3$ ,  $k_F\xi_c = 40/\sqrt{3}$ .



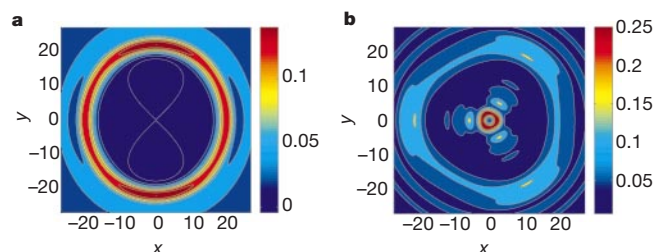
quasiparticle spectrum for small angular momentum quantum numbers  $\mu$  is  $\epsilon_\mu = \mu\Delta/|\mathbf{k}_\perp|\xi$  (ref. 7), where  $\Delta$  is the superconducting gap far from the vortex axis,  $\xi$  is the coherence length,  $|\mathbf{k}_\perp|$  is the absolute value of the wavevector in the plane perpendicular to the vortex, and  $\mu$  is half an odd integer. This is the so-called anomalous branch of the quasiparticle energy, which, as function of  $\mu$ , varies from  $-\Delta$  to  $\Delta$ , crossing zero as the impact parameter  $b = \mu/|\mathbf{k}_\perp|$  of the particle in the core varies from  $-\infty$  to  $\infty$ . For  $m > 1$ , the number of anomalous energy branches (per spin) crossing the Fermi level is equal to  $m$  (refs 8–10). If a vortex carries an odd number of flux quanta, one of these energy branches crosses the Fermi level at zero angular momentum  $\mu = 0$  corresponding to  $b = 0$  and, thus, the peak in the quasiparticle DOS at the vortex centre appears. A vortex with an even number  $m$  does not have such an energy branch, and the DOS peak at the vortex centre disappears. All anomalous branches in this latter case cross the Fermi level at finite impact parameter  $b \approx r_c$ , where  $r_c$  is the core radius. Making use of the quasiclassical approach, we obtain the quasiparticle energy spectrum as a function of the angular momentum quantum number  $\mu$  and find the corresponding spatial distribution of the DOS  $N(\epsilon, \mathbf{r})$ . Shown in Fig. 2 is a thermodynamically averaged local DOS  $g(\mathbf{r})$  normalized by its normal state value  $g_N$ , for  $m = 2$  (Fig. 2a) and  $m = 3$  (Fig. 2b):

$$g(\mathbf{r}) = \int_{-\infty}^{\infty} \frac{N(\epsilon, \mathbf{r}) d\epsilon}{4T \cosh^2(\epsilon/2T)} \quad (1)$$

$g(\mathbf{r})$  is directly related to a zero-bias local tunnelling conductance between two reservoirs, and can be probed by scanning tunnelling spectroscopy<sup>11</sup>.

Now we discuss DOS in the vortex molecule state. Treating splitting of a multi-quanta vortex into a small molecule with size  $a \leq \xi$  as a perturbation, we observe that as soon as single-quanta vortices start to separate, each ring of the maximal DOS of a multi-quanta vortex breaks into  $m$  peaks. With an increase in the size of the molecule and, accordingly, in vortex spacing, some of the DOS peaks merge, and finally only  $m$  peaks at the centres of individual vortices survive. Typical spatial distributions of the corresponding thermodynamically averaged local DOS derived for two particular cases of  $2\Phi_0$  and  $3\Phi_0$  vortex molecules ( $m = 2$  and  $m = 3$ ) are shown in Fig. 3.

In the small samples with radius  $R$  comparable to the coherence length, the electronic structure is strongly affected by the edge electronic states. The finite magnetic field suppresses the order parameter near the disk edge, and creates a potential well for quasiparticles. Bound quasiparticle states form, owing to confinement by normal quasiparticle reflection at the disk edge and by Andreev reflection from the boundary of the classically impenetrable region. We derive the spectrum as a function of the angular momentum and magnetic field using the procedure analogous to that of refs 12, 13. Shown in Fig. 4 are the energy branches for a Meissner state ( $m = 0$ ) as functions of dimensionless impact parameter  $\tilde{b} = \mu/(k_\perp R)$ . At  $\mu \approx |k_\perp| R$  energy branches have



**Figure 3** Spatial distribution of the thermodynamically averaged local DOS  $g(\mathbf{r})$  for a vortex molecule. **a**,  $m = 2$ ; **b**,  $m = 3$ . The  $x, y$  coordinates are measured in  $1/k_F$ ,  $T \approx \Delta/k_F \xi$ . For  $m = 2$ ,  $k_F r_c = 20\sqrt{2}$ ; for  $m = 3$ ,  $k_F r_c = 40/\sqrt{3}$ .

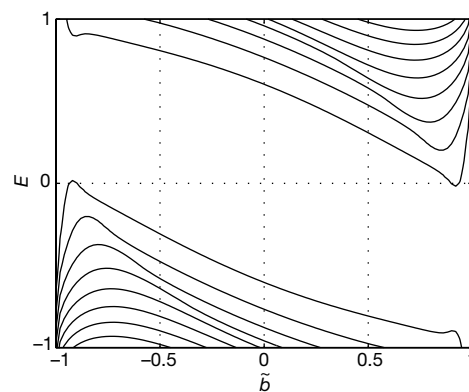
minima, resulting in the sequence of discontinuities (step-like structure) in the energy and magnetic-field dependences of DOS. The appearance of a multi-quanta vortex at the disk centre gives rise to switching between the energy branches with different  $m$ : each subsequent flux quantum entering the disk reduces the depth of the edge potential well for quasiparticle excitations, and hence shifts the localized levels to higher energies. A further increase of the magnetic field increases the well depth (until the next flux quantum hops in). Therefore the DOS will show oscillations with the period  $\delta H \approx \Phi_0/R^2$ . The amplitude of oscillations increases with the decrease in the disk size, and is observable only in mesoscopic samples.

Now we turn to phase-coherent ballistic transport through the quasiparticle vortex states. Consider a mesoscopic superconducting disk squeezed between two contacts placed in the very centre of the disk (see Fig. 1: such contacts can be realized, for example, by STM tips pressed into disk surfaces). As the size of a few-fluxoid sample becomes less than the dephasing length, new physical processes come into play. Even in the case of a gapped quasiparticle spectrum and at the zero-temperature limit, single-particle ballistic transport through sufficiently thin disks can be mediated by quasiparticles with wavefunctions decaying in the current direction. Every flux quantum that enters the thin disk and joins either the vortex molecule or the multi-quanta giant vortex opens an additional channel for single-particle tunnelling. Thus the subsequent trapping of flux quanta results in a step-like increase in the disk conductance.

In a macroscopic superconducting system, the core states do not contribute to the electro-conductivity along the vortex lines, because the quasiparticle channels are shunted by the condensate: the normal current injected at the surface into the centre of the core converts to supercurrent over a certain length scale—which is less than the sample size. In the mesoscopic sample, the current does not have time to go outside the core, and one expects core states to dominate conductance. Yet the quasiparticle states still decay into the sample owing to multiple Andreev reflections from the boundaries of the vortex core. Our first step in the derivation of the quasiparticle conductivity is then the estimation of the decay length for each quasiparticle mode, and finding the corresponding fraction of the conducting channels contributing effectively to conductance. To this end, we look for the waveguide-like solution to the Bogolubov–de Gennes equations, and find the inverse decay length as:

$$\gamma_n \approx \frac{n \ln \Lambda}{k_\parallel \xi^2}, \quad (2)$$

where  $n$  is an integer,  $\Lambda \approx k_F \xi$ , and  $k_\parallel$  is the wavevector component



**Figure 4** Energy spectrum for electronic states at the sample edge. Parameters used in these calculations are:  $H = H^* = \phi_0/(\pi R \xi)$ ,  $R = 15\xi$ ,  $|\mathbf{k}_\perp|/k_F = 1$  (see text for details).  $\tilde{b} = \mu/(|\mathbf{k}_\perp| R)$  is the dimensionless impact parameter.

parallel to the vortex axis.  $k_F$  is the Fermi momentum. At small  $n$ , the decay length becomes much larger than the coherence length.

We now consider the ballistic conductance of the vortex of the length  $L$ . Consider the superconducting disk with the vortex as placed between the two normal metal reservoirs. Each quasiparticle mode contribution into the conductance is proportional to  $(e^2/h) \exp(-2\gamma_n L)$ . Then the total conductance due to single-particle tunnelling processes becomes:

$$G \propto \frac{e^2}{h} T_b \sum_{\mu} \sum_{n=n_{\min}}^{n_{\max}} \exp(-2\gamma_n L) \quad (3)$$

where  $n_{\min} \approx \mu$ ,  $n_{\max} \approx k_F \xi$ . We have assumed here the barrier transmission probability  $T_b$  to be the same for all channels (in a more detailed model, this coefficient should depend on the quantum numbers). In the case of interest we focus on,  $\xi < L < k_F \xi^2$ . Replacing sums over  $\mu$  and  $n$  by the integrals, we arrive at the final estimate:

$$G \propto \frac{e^2}{h} T_b (k_F \xi)^2 \frac{\xi^2}{L^2} \quad (4)$$

For small samples of the size of several coherence lengths, single-particle conductance compares favourably with the two-particle Andreev contribution ( $\propto T_b^2$ ). At large distances,  $L > k_F \xi^2$ , the  $1/L^2$  power-law decay of the conductance crosses over to the exponential reduction  $\propto \exp[-L/(k_F \xi^2)]$ , and in the long samples the single-particle current through the vortex channels is shunted by the supercurrent through the borders of the vortex core.

For the schematic experimental design shown in Fig. 1, we see that there is a notable dependence of the ballistic conductance behaviour on the magnetic field that exists at the area of the contact. For large-area contacts, all the additional channels—corresponding to new maxima of the DOS—that appear as flux quanta subsequently enter the disk, contributing to the conductivity. Hence such a ‘large-contact-area device’ will exhibit a step-like growth of conductance with increasing magnetic field (see Fig. 1). On the contrary, a point contact with an area much smaller than the area of the vortex core feels the density of states just at its location, and thus shows a strong dependence of the transmission probability  $T_b$  on  $\mu$  and on contact position. Placing the point contact strictly at the centre of the disk, we observe either nearly zero (for even  $m$ , where the central spot in the DOS is absent) or finite alternating conductance for odd  $m$ , where single-particle tunnelling occurs via the central peak (see Fig. 1). This alternating conductance realizes a quantum switch, with magnetic field playing the role of gate voltage. Generally, the magnetic-field dependence of conductance will exhibit a mixture of the above step-like and alternating behaviours.

The ballistic contribution to the conductance dominates the quasiparticle transport at rather low temperatures,  $T < \Delta \xi/L$ ; in all other cases, the conductance will be determined by the thermodynamically averaged DOS. Thus, experimental realization of ballistic quantum switch regime requires low temperatures—and pure, thin samples. □

Received 16 August; accepted 7 November 2001.

- Boato, G. *et al.* Direct evidence for quantized flux threads in type II superconductors. *Solid State Commun.* **3**, 173–176 (1965).
- McLachlan, D. S. Quantum oscillations and the order of the phase charge in a low  $k$  type II superconducting microcylinder. *Solid State Commun.* **8**, 1589–1593 (1970).
- Geim, A. K. *et al.* Fine structure in magnetization of individual fluxoid states. *Phys. Rev. Lett.* **85**, 1528–1531 (2000).
- Schweigert, V. A., Peeters, F. M. & Deo, P. S. Vortex phase diagram for mesoscopic superconducting disks. *Phys. Rev. Lett.* **81**, 2783–2786 (1998).
- Chibotaru, L. F. *et al.* Symmetry-induced formation of antivortices in mesoscopic superconductors. *Nature* **408**, 833–835 (2000).
- Giaever, I. in *Tunneling Phenomena in Solids* (eds Burstein, E. & Lundqvist, S.) 255–271 (Plenum, New York, 1969).
- Caroli, C., de Gennes, P. G. & Matricon, J. Bound fermion states on a vortex line in a type II superconductor. *Phys. Lett.* **9**, 307–309 (1964).
- Volovik, G. E. Vortex motion in Fermi superfluids and the Callan-Harvey effect. *JETP Lett.* **57**, 244–246 (1993).

- Tanaka, Y. *et al.* Energy spectrum of the quasiparticle in a quantum dot formed by a superconducting pair potential under a magnetic field. *Solid State Commun.* **85**, 321–326 (1993).
- Virtanen, S. M. M. & Salomaa, M. M. Multiquantum vortices in superconductors: Electronic and scanning tunneling microscopy spectra. *Phys. Rev. B* **60**, 14581–14584 (1999).
- Hess, H. F. *et al.* Scanning-tunneling-microscope observation of the Abrikosov flux lattice and the density of states near and inside a fluxoid. *Phys. Rev. Lett.* **62**, 214–217 (1989).
- Pincus, P. Magnetic-field-induced surface states in a pure type-I superconductor. *Phys. Rev. B* **158**, 346–353 (1967).
- Azbel, M. Ya. & Blank, A. Ya. Magnetic surface levels in superconductors. *JETP Lett.* **10**, 32–35 (1969).

## Acknowledgements

We thank G. Crabtree, G. Karapetrov, A. Koshelev, V. Moshchalkov, D. A. Ryzhov, S. V. Sharov, I. A. Shereshevsky and I. D. Tokman for discussions. This work was supported by the US DOE Office of Science, a NATO Collaborative Linkage Grant, and by the Russian Foundation for Fundamental Research.

Correspondence and requests for materials should be addressed to V.M.V. (e-mail: vinokur@msd.anl.gov).

## A robust DNA mechanical device controlled by hybridization topology

Hao Yan, Xiaoping Zhang, Zhiyong Shen & Nadrian C. Seeman

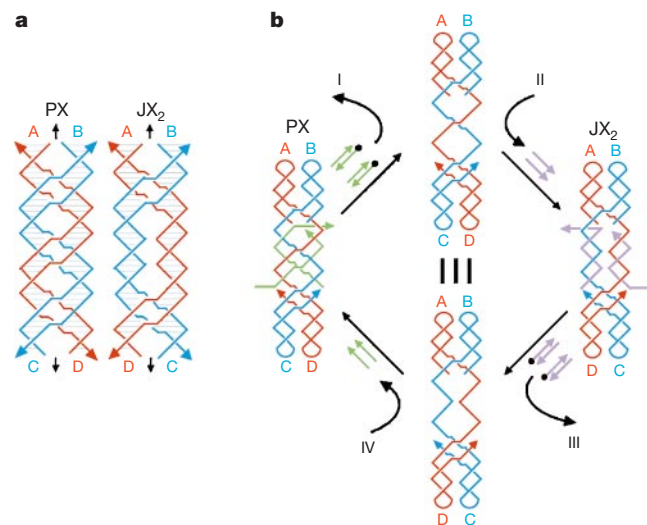
Department of Chemistry, New York University, New York, New York 10003, USA

Controlled mechanical movement in molecular-scale devices has been realized in a variety of systems—catenanes and rotaxanes<sup>1–3</sup>, chiroptical molecular switches<sup>4</sup>, molecular ratchets<sup>5</sup> and DNA<sup>6</sup>—by exploiting conformational changes triggered by changes in redox potential or temperature, reversible binding of small molecules or ions, or irradiation. The incorporation of such devices into arrays<sup>7,8</sup> could in principle lead to complex structural states suitable for nanorobotic applications, provided that individual devices can be addressed separately. But because the triggers commonly used tend to act equally on all the devices that are present, they will need to be localized very tightly. This could be readily achieved with devices that are controlled individually by separate and device-specific reagents. A trigger mechanism that allows such specific control is the reversible binding of DNA strands, thereby ‘fuelling’ conformational changes in a DNA machine<sup>9</sup>. Here we improve upon the initial prototype system that uses this mechanism but generates by-products<sup>9</sup>, by demonstrating a robust sequence-dependent rotary DNA device operating in a four-step cycle. We show that DNA strands control and fuel our device cycle by inducing the interconversion between two robust topological motifs, paranemic crossover (PX) DNA<sup>10,11</sup> and its topoisomer JX<sub>2</sub> DNA, in which one strand end is rotated relative to the other by 180°. We expect that a wide range of analogous yet distinct rotary devices can be created by changing the control strands and the device sequences to which they bind.

PX DNA (Fig. 1a, left) is a four-stranded molecule in which two parallel double helices are joined by reciprocal exchange (crossing over) of strands at every point where the strands come together<sup>10,11</sup>. JX<sub>2</sub> (Fig. 1a, right) is a topoisomer of PX DNA that contains two adjacent sites where backbones juxtapose without crossing over. The colour coding of the strands and labels in Fig. 1 indicates that the top ends, A and B, are the same in both molecules, but the bottom ends, C and D, are rotated 180°. This rotation is the basis for the operation of the device. We have used strand replacement<sup>9</sup> to interconvert the PX and JX<sub>2</sub> motifs.

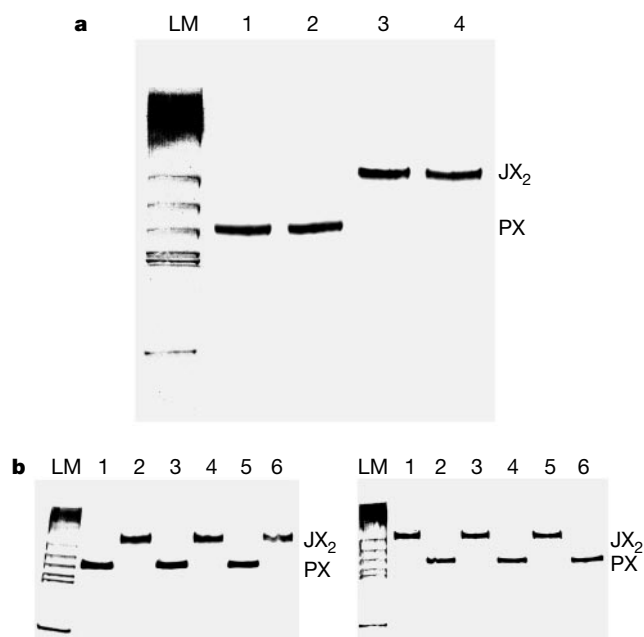
In the device that we have constructed, one strand of each of the blue and red strand pairs is broken into three strands. The principles of operation are illustrated in Fig. 1b, where the red and blue strands of opposite polarity are shown connected by hairpin loops. Thus, the PX molecule shown there consists of one red strand, one blue strand, and two green strands, termed the 'set' strands, because they set the state of the device to be in the PX conformation; similarly, the JX<sub>2</sub> molecule has purple set strands. The set strand associated with the red strand has a 5' single-stranded extension, and the set strand associated with the blue strand has a similar 3' extension. Extensions like these can<sup>9</sup> be used to initiate branch migration that leads to removal of the strand from the branched motif, because it is paired with a complementary strand along its entire length. Thus, a complement to the entire length of the set strand (termed a 'fuel' strand) will pair with it in preference to the partially paired set strand in the PX (or JX<sub>2</sub>) motif.

Process I (Fig. 1b) shows the addition of fuel strand complements to the two green set strands of the PX device, producing the unstructured intermediate at the top of the drawing. Process II shows the addition of pale-purple set strands that convert the intermediate to the JX<sub>2</sub> conformation. Process III shows the addition of fuel strands that convert the JX<sub>2</sub> molecule to the unstructured intermediate, and process IV shows the addition of the green set strands to produce the PX conformation again. Alternation between a paired structure and a partially unpaired structure analogous to this intermediate characterized the action of the system of ref. 9. In the device described here, the four-step cycle leads to two robust end points, the PX state and the JX<sub>2</sub> state.



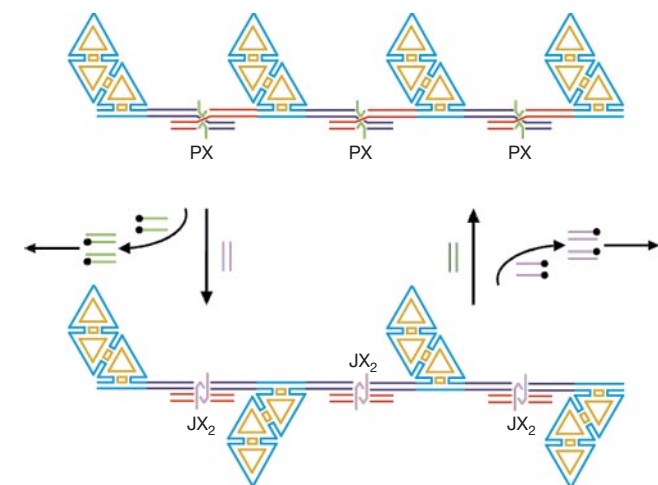
**Figure 1** Schematic drawings of the device. **a**, The PX and JX<sub>2</sub> motifs. The PX motif, postulated to be involved in genetic recombination<sup>11</sup>, consists of two helical domains formed by four strands that flank a central dyad axis (indicated by the vertical black arrows). Two strands are drawn in red and two in blue, where the arrowheads indicate the 3' ends of the strands. The Watson-Crick base pairing in which every nucleotide participates is indicated by the thin horizontal lines within the two double helical domains. The same conventions apply to the JX<sub>2</sub> domain, which lacks two crossovers in the middle. The letters A, B, C and D, along with the colour coding, show that the bottom of the JX<sub>2</sub> motif (C and D) are rotated 180° relative to the PX motif. **b**, Principles of device operation. On the left is a PX molecule. The green set strands are removed by the addition of biotinylated green 'fuel' strands (biotin indicated by black circles) in process I. The unstructured intermediate is converted to the JX<sub>2</sub> motif by the addition of the pale-purple set strands in process II. The JX<sub>2</sub> molecule is converted to the unstructured intermediate by the addition of biotinylated pale-purple 'fuel' strands in process III. The identity of this intermediate and the one above it is indicated by the identity symbol between them. The cycle is completed by the addition of green set strands in process IV, restoring the PX device.

To demonstrate the operation of a robust molecular mechanical device, it is necessary to show both the uniform behaviour of the bulk material, and also to visualize the structural transformations of selected molecules. Figure 2a illustrates the formation and inter-conversion of both PX and JX<sub>2</sub> DNA by non-denaturing gel electrophoresis. All experiments are performed in a buffer containing 40 mM Tris-HCl, pH 8.0, 20 mM acetic acid, 2 mM EDTA and 12.5 mM magnesium acetate; sequences of all molecules used are in the Supplementary Information. The absence of species other than the PX or JX<sub>2</sub> molecules (for example, the dimers noted by Yurke *et al.*<sup>9</sup>, or potential dissociation products) attests to the robustness of the device in bulk. Figure 2b illustrates the cycling of the device between the PX and JX<sub>2</sub> states. The right portion of Fig. 2b shows five steps of operation, beginning in the JX<sub>2</sub> state; the left portion shows five steps that begin from the PX state. The intermediate structure is stable at the highest temperatures (20 °C) to which the device has been subjected in this experiment (see Supplementary Information).



**Figure 2** Gel evidence for the operation of the device. **a**, The components of the device in operation. This is a 14% non-denaturing polyacrylamide gel, run at 20 °C and stained with stains-all dye (Sigma E-7762). The lane LM contains linear length markers derived from *Hae*III digestion of pBR322. Device strand sequences have been designed using the program SEQUIN<sup>13</sup>, synthesized by routine phosphoramidite techniques<sup>14</sup> and gel purified. Strands are hybridized at 90 °C (5 min), 65 °C (15 min), 45 °C (30 min), 37 °C (20 min) and 20 °C (30 min). Lane 1 contains the device [1 μM] assembled with PX set strands and lane 4 contains the device [1 μM] assembled with JX<sub>2</sub> set strands. Gel mobilities differ because the PX device is likely to have a more compact time-averaged structure than the JX<sub>2</sub> device<sup>15</sup>. Lane 2 contains the products of removing the JX<sub>2</sub> set strands from the material in lane 4 and replacing them with set strands corresponding to the PX conformation. Likewise, lane 3 contains the products of removing the PX set strands from the material in lane 1 and replacing them with those corresponding to the JX<sub>2</sub> conformation. Note the absence of extraneous products in lanes 2 and 3, indicating the robustness of these transformations. **b**, Cycling the device. The two portions of this panel show the cycling of the device through 5 steps. Lane 1 is the initial conformation (JX<sub>2</sub> right and PX left), and lanes 2 to 6, respectively, show alternating transformations to the other state. Fuel strands were added to the preformed PX or JX<sub>2</sub> at 20 °C and kept at 20 °C for 60 min; the mixture was treated with streptavidin beads at 20 °C for 30 min to remove the set-strand/fuel-strand duplexes. After removing the set strands of PX or JX<sub>2</sub>, the set strands of JX<sub>2</sub> or PX molecules were added to the solution and kept at 20 °C for 60 min to establish the device conformation. The addition of fuel strands, followed by set strands, was then repeated three times.



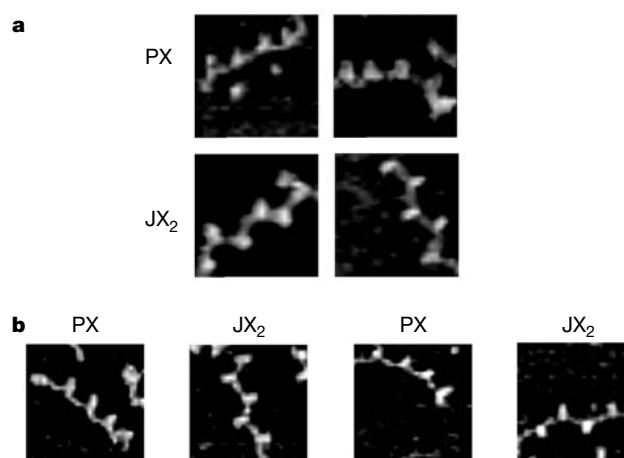


**Figure 3** A highly simplified representation of the system used. It consists of a one-dimensional array of half-hexagons (light-blue outer strands, ochre inner strands) joined by the device. Each half-hexagon consists of three edge-sharing DNA triangles<sup>16</sup> whose edges are three turns long; the edge-sharing structure is a DNA double crossover molecule<sup>12</sup>, which also attaches the half-hexagon to the linear components of the array. The colour coding in the vicinity of the device has been retained from Fig. 1, with green set strands in the PX state, pale-purple set strands in the JX<sub>2</sub> state and constant strands in red and blue; the actual strand structure is both more complex and larger than the structure in Fig. 2, and is shown in detail with its sequence in the Supplementary Information. There are 39 nucleotides between the first device crossover point and the nearest triangle crossover point, a number that was determined empirically to give the most nearly planar structure, although it represents four turns of DNA. In the upper molecule, all of the half-hexagons are aligned pointing in the same direction (*cis*), whereas they point in opposite directions in the bottom molecule (*trans*). Biotinylated fuel strands (with black filled circles) are shown removing set strands in both parts of the cycle. Note that relative to the half-hexagon on the left, the third one has rotated 360° and the rightmost one has rotated 540°.

Altered gel mobilities do not guarantee that the construct undergoes the designated structural transformation. We demonstrate this aspect of the device with the system shown in Fig. 3. We have constructed half-hexagon markers via edge-sharing between three triangles; each of the shared edges is a DNA double crossover molecule<sup>12</sup>. The half-hexagons are connected into one-dimensional oligomeric arrays by linkage through extensions that include PX-JX<sub>2</sub> devices. Figure 3a shows that if the devices are all in the PX state, the half-hexagons have a *cis* arrangement, where they all point in the same direction. However, when the devices are all in the JX<sub>2</sub> state, the half-hexagons form a zigzag *trans* structure.

Figure 4 illustrates the operation of the device, using atomic force microscopy (AFM) images. Figure 4a shows two examples for the *cis* and *trans* systems of Fig. 3, but containing links that are fixed to be PX or JX<sub>2</sub> molecules, rather than PX-JX<sub>2</sub> devices. The PX state contains a series of half-hexagons extended in a parallel direction, much like the extended fingers of a hand. By contrast, the JX<sub>2</sub> state is characterized by a zigzag arrangement of the half-hexagon extensions. Figure 4b illustrates the operation of the device by displaying representative molecules sampled from solutions expected to contain successively (left to right) PX, JX<sub>2</sub>, PX and JX<sub>2</sub> states, as the system is cycled. The PX molecules have their half-hexagon markers aligned in a *cis* arrangement, whereas the markers in the JX<sub>2</sub> molecules are all *trans*. Thus, the system operates as designed, both in bulk and in individual cases. The intermediate state produces a single band on a gel, but it is not well-structured when examined by AFM (data not shown).

We have shown that a rotary nanomechanical device is capable of being cycled by the addition of strands that direct its structure; we



**Figure 4** Evidence for the operation of the device, obtained from atomic force microscopy (AFM). All initial species are produced by heating their constituent single strands in boiling water and then cooling in a styrofoam box over a period of 2 days. The one-dimensional arrays of these half-hexagon-plus-device units cohere by way of 8-nucleotide sticky ends. Non-denaturing gels demonstrate the resistance of this sticky end to disruption at 45 °C, where these conversions were performed (see Supplementary Information).

**a**, Control images. These images contain control molecules, not devices, that are constrained to be in the PX or JX<sub>2</sub> motifs. AFM samples are prepared by placing 1 μl of solution on a piece of freshly cleaved mica (Ted Pella, Inc.), blowing it dry, and washing several times with distilled water. Images are obtained in isopropanol by scanning with a Nanoscope II in contact mode. The upper panels show PX linear arrays in a *cis* arrangement, and the lower panels show JX<sub>2</sub> linear arrays in a *trans* arrangement.

**b**, Cycling of the device. This panel shows three steps of the operation of the device, sampling aliquots from each cycle. The system originates in the PX state, and is then converted (left to right) to the JX<sub>2</sub> state, back to PX, and then back to JX<sub>2</sub>. The PX linear arrays are clearly in the *cis* arrangement, and the JX<sub>2</sub> linear arrays are clearly in the *trans* arrangement. All images show an area 200 nm × 200 nm.

describe the system as robust, because both end points—the PX state and the JX<sub>2</sub> state—are well-defined structures that lack single-stranded regions in structural roles. The extent of motion produced within the rotary device itself will be a function of the distance from its midline, ranging from about 0.4 to 4 nm; however, motions as large as 35 nm have been achieved with the half-hexagon array. Multiple species could be obtained by changing the set strands and the sequences to which they bind. If *N* different device species of this type can be incorporated into two-dimensional<sup>7,8</sup> or three-dimensional crystalline arrays, 2<sup>*N*</sup> different structural states will be available to the system. Multiple robust states of this sort are necessary for an effective nanorobotics, so that a diversity of shapes can be programmed. Further experimentation will be required to establish that this DNA device can transmit a force capable of performing useful work on other chemical species. □

Received 26 April; accepted 8 November 2001.

- Pease, A. R. *et al.* Switching devices based on interlocked molecules. *Acc. Chem. Res.* **34**, 433–444 (2001).
- Jimenez, M. C., Dietrich-Buchecker, C. & Sauvage, J.-P. Towards synthetic molecular muscles: contraction and stretching of a linear-rotaxane dimer. *Angew. Chem. Int. Edn Engl.* **39**, 3284–3287 (2000).
- Brouwer, A. M. *et al.* Photoinduction of fast, reversible translational motion in a hydrogen-bonded molecular shuttle. *Science* **291**, 2124–2128 (2001).
- Koumura, N., Zijlstra, R. W. J., van Delden, R. A., Harada, N. & Feringa, B. L. Light-driven monodirectional molecular rotor. *Nature* **401**, 152–155 (1999).
- Kelly, T. R., De Silva, H. & Silva, R. A. Unidirectional rotary motion in a molecular system. *Nature* **401**, 150–152 (1999).
- Mao, C., Sun, W., Shen, Z. & Seeman, N. C. A DNA nanomechanical device based on the B-Z transition. *Nature* **397**, 144–146 (1999).
- Winfree, E., Liu, F., Wenzler, L. A. & Seeman, N. C. Design and self-assembly of two-dimensional DNA crystals. *Nature* **394**, 539–544 (1998).
- LaBean, T. H. *et al.* The construction, analysis, ligation and self-assembly of DNA triple crossover complexes. *J. Am. Chem. Soc.* **122**, 1848–1860 (2000).

9. Yurke, B., Turberfield, A. J., Mills, A. P. Jr, Simmel, F. C. & Neumann, J. L. A DNA-fuelled molecular machine made of DNA. *Nature* **406**, 605–608 (2000).
10. Seeman, N. C. DNA nicks and nodes and nanotechnology. *Nano Lett.* **1**, 22–26 (2001).
11. Shen, Z. *DNA Polycrossover Molecules and their Applications in Homology Recognition*. Thesis, New York Univ. (1999).
12. Fu, T.-J. & Seeman, N. C. DNA double crossover structures. *Biochemistry* **32**, 3211–3220 (1993).
13. Seeman, N. C. De novo design of sequences for nucleic acid structure engineering. *J. Biomol. Struct. Dyn.* **8**, 573–581 (1990).
14. Caruthers, M. H. Gene synthesis machines: DNA chemistry and its uses. *Science* **230**, 281–285 (1985).
15. Sun, W., Mao, C., Iwasaki, H., Kemper, B. & Seeman, N. C. No braiding of Holliday junctions in positively supercoiled DNA molecules. *J. Mol. Biol.* **294**, 683–699 (1999).
16. Yang, X., Wenzler, L. A., Qi, J., Li, X. & Seeman, N. C. Ligation of DNA triangles containing double crossover molecules. *J. Am. Chem. Soc.* **120**, 9779–9786 (1998).

**Supplementary Information** accompanies the paper on Nature's website (<http://www.nature.com>).

## Acknowledgements

We thank R. Sha for discussions. This work was supported by the Office of Naval Research, the National Institute of General Medical Sciences, the National Science Foundation/DARPA, the Information Directorate of the Air Force Research Laboratory located at Rome, New York, and the National Science Foundation.

Correspondence and requests for materials should be addressed to N.C.S. (e-mail: ned.seeman@nyu.edu).

# Evolutionary speed limits inferred from the fossil record

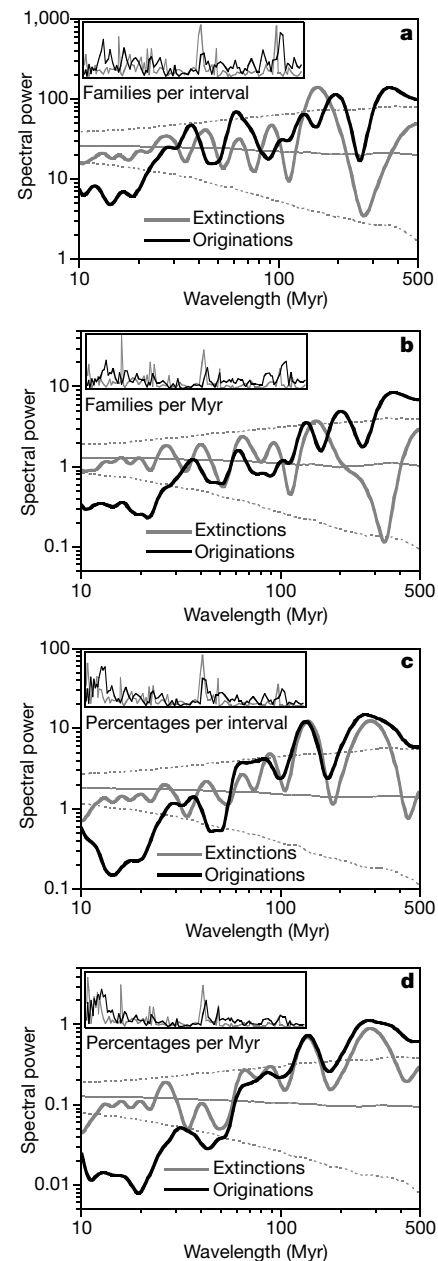
James W. Kirchner

Department of Earth and Planetary Science, University of California, Berkeley, California 94720-4767, USA

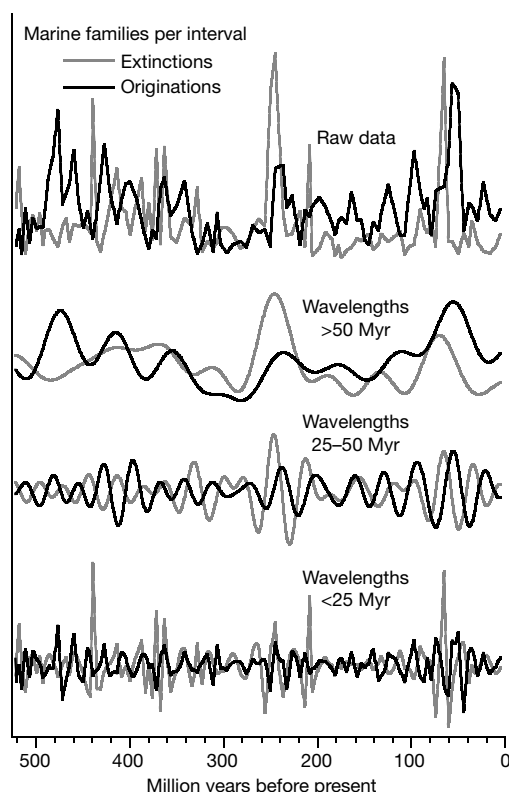
The dynamics of extinction and diversification determine the long-term effects of extinction episodes<sup>1</sup>. If rapid bursts of extinction are offset by equally rapid bursts of diversification, their biodiversity consequences will be transient. But if diversification rates cannot accelerate rapidly enough, pulses of extinction will lead to long-lasting depletion of biodiversity<sup>2,3</sup>. Here I use spectral analysis of the fossil record to test whether diversification rates can accelerate as much as extinction rates, over both short and long spans of geological time. I show that although the long-wavelength variability of diversification rates equals or exceeds that of extinctions, diversification rates are markedly less variable than extinction rates at wavelengths shorter than roughly 25 million years. This implies that there are intrinsic speed limits that constrain how rapidly diversification rates can accelerate in response to pulses of extinction.

To measure how fossil extinction and origination rates fluctuate across different timescales, I used spectral analysis methods<sup>4,5</sup> specially designed for unevenly spaced data such as the fossil record. Spectral analysis decomposes a time series into its component frequencies and measures each of their amplitudes (expressed as their square, the spectral power). My source data are Sepkoski's compilations of Phanerozoic fossil marine animal families<sup>6</sup> and genera<sup>7</sup>. Because the suitability of different extinction and origination metrics is sometimes disputed<sup>8,9</sup>, I analysed all four that are in common use: (1) counts of first or last occurrences per stratigraphic interval; (2) counts per million years (counts in each stratigraphic interval, divided by interval length); (3) percentages per interval (counts divided by total diversity); and (4) percentages per million years (counts divided by total diversity and interval length). This yields a total of eight data sets (two taxonomic levels times four metrics) for both extinctions and originations.

Spectral analysis shows that across all eight data sets, extinctions and originations have fundamentally different scaling behaviour (Fig. 1; and Supplementary Information). From the shortest resolvable timescale (~10 Myr, twice the average sampling interval) to the longest (~500 Myr, the length of the data set), the spectral power of extinction rates usually lies within the confidence limits for white noise, as estimated by randomly shuffling and re-analysing the



**Figure 1** Fourier power spectra for extinction and origination rates of fossil marine families. Four different metrics were used: families per stratigraphic interval (**a**) and per million years (**b**), and percentages of total diversity per stratigraphic interval (**c**) and per million years (**d**). Insets show the raw time series. Spectra are smoothed over a window spanning 5% of the width of each plot. Fine lines are the median and 90% confidence limits for spectral power of random white noise at each wavelength, estimated by repeating the same analysis on 10,000 random re-shuffles of the original extinction data. Spectral power is the square of amplitude at each wavelength; thus, the relative placement of extinction and origination spectra is not arbitrary. Origination rates have markedly lower spectral power than extinction rates at wavelengths less than roughly 25 Myr, indicating that they are much less variable over short timescales. Spectra for genera are similar (see Supplementary Information).



**Figure 2** Fossil extinction and origination rates, separated into long-wavelength (> 50 Myr), intermediate-wavelength (25–50 Myr) and short-wavelength (< 25 Myr) components by Fourier filtering<sup>26</sup>. Long-wavelength variations in origination and extinction rates are similar in size, but short-wavelength fluctuations in extinction rates are

larger and more abrupt than those in origination rates. Counts are shown of fossil marine families originating or going extinct in each stratigraphic interval; plots for other extinction and origination metrics are similar (see Supplementary Information).

original data. (Extinction rates that are expressed as percentages of total diversity have greater spectral power than expected by chance at wavelengths greater than 100 Myr, but this may reflect long-wavelength patterns resulting from changes in diversity through time, rather than long-term patterns in extinction rates *per se*.)

Compared with extinction rates, origination rates (Fig. 1, black lines) have equal or greater spectral power at long wavelengths (> 100 Myr), but much lower spectral power at short wavelengths (< 25 Myr). That is, origination rates fluctuate less than extinction rates over short timescales, even though over long timescales they fluctuate as much as, or more than, extinction rates. Fossil origination rates are known to vary less than extinction rates<sup>9,10</sup>. My analysis shows that this results from the markedly lower short-term variability of originations, because on longer timescales origination rates vary at least as much as extinction rates do.

The average spectral power in the 10–25-Myr wavelength band is

3.4 times lower for origination rates than for extinction rates (averaged across the eight data sets), implying that short-wavelength fluctuations are 1.8 times smaller in originations than in extinctions. This difference in fluctuation scaling can also be seen in the statistics of the time series themselves. Across the eight data sets analysed here, origination rates are 30% higher, on average, than extinction rates (consistent with the overall growth in diversity through time), but are a factor of 1.6 less variable between adjacent stratigraphic intervals (Table 1). Furthermore, the greatest increases in origination rates from one stratigraphic interval to the next are a factor of 2.3 smaller, on average across the eight data sets, than the greatest increases in extinction rates (Table 1). Thus, by several different measures, origination rates show substantially less short-term variability than extinction rates.

The difference in scaling behaviour between originations and extinctions can be visualized by using Fourier filtering to separate

**Table 1 Fossil origination and extinction rate statistics**

| Taxonomic level and metric | Average rates |       |       | Rate changes (r.m.s.)* |       |       | Maximum increases† |       |       |
|----------------------------|---------------|-------|-------|------------------------|-------|-------|--------------------|-------|-------|
|                            | Ext.          | Orig. | Ratio | Ext.                   | Orig. | Ratio | Ext.               | Orig. | Ratio |
| Families                   |               |       |       |                        |       |       |                    |       |       |
| per interval               | 25.1          | 34.8  | 0.72  | 34.5                   | 22.8  | 1.5   | 130                | 54    | 2.4   |
| per Myr                    | 5.3           | 7.1   | 0.74  | 39.7                   | 20.8  | 1.7   | 32.8               | 8.9   | 3.7   |
| % per interval             | 6.0           | 7.8   | 0.78  | 37.7                   | 29.7  | 1.3   | 19.0               | 18.8  | 1.0   |
| % per Myr                  | 1.4           | 1.7   | 0.81  | 9.1                    | 7.7   | 1.4   | 6.4                | 4.1   | 1.6   |
| Genera                     |               |       |       |                        |       |       |                    |       |       |
| per interval               | 149           | 197   | 0.76  | 1,049                  | 787   | 1.6   | 900                | 396   | 2.3   |
| per Myr                    | 30.4          | 39.2  | 0.78  | 228.8                  | 145.4 | 1.7   | 192.0              | 77.3  | 2.5   |
| % per interval             | 13.4          | 16.0  | 0.84  | 60.8                   | 42.6  | 1.6   | 38.8               | 25.0  | 1.6   |
| % per Myr                  | 2.9           | 3.4   | 0.86  | 16.9                   | 12.8  | 1.6   | 14.1               | 6.0   | 2.4   |

\* Root mean-square average of changes in rates between adjacent stratigraphic intervals.

† Maximum increase in rates from one interval to the next. Ext., extinctions; Orig., originations.



the original time series' fluctuations into specified ranges of wavelengths (Fig. 2; and Supplementary Information). Fluctuations in originations and extinctions are generally similar in amplitude (although they often do not coincide in time) at wavelengths longer than 25 Myr (Fig. 2). But at wavelengths shorter than 25 Myr, extinction rates show large, abrupt fluctuations that are not observed in origination rates.

Whereas the power spectra of extinction rates generally lie within the confidence limits for random white noise, the power spectra of origination rates generally lie well outside those confidence limits at wavelengths shorter than about 25 Myr (Fig. 1; and Supplementary Information). The log-log slopes of the extinction power spectra are generally small, and often statistically indistinguishable from white noise, whereas the log-log slopes of the origination power spectra are markedly steeper, and statistically highly significant (Table 2). The coarseness of the stratigraphic timescale makes it impossible to measure accurately fluctuations on wavelengths shorter than about 10 Myr (see Methods), but if the scaling trends shown in Fig. 1 were to continue at shorter wavelengths, they would imply that the discrepancy between short-term fluctuations in originations and extinctions should grow as the temporal resolution becomes finer.

Why are diversification rates less variable than extinction rates over short timescales? Uneven preservation and sampling through geological time<sup>11,12</sup> should affect measurements of origination and extinction rates to similar degrees, and thus are unlikely to account for origination rates' markedly smaller short-term variability. Instead, this phenomenon implies either that the processes regulating originations have more inertia than those driving extinctions, or that origination events tend to be diverse and local, whereas extinctions (particularly mass extinctions) tend to be coherent and global<sup>9,13,14</sup>. If extinctions are driven primarily by external shocks<sup>2,15</sup> they are unlikely to be intrinsically rate-limited, whereas originations of new taxa may be limited by sparse opportunities for reproductive isolation and thus speciation. Alternatively, the process of diversification may be inhibited by harsh environmental conditions during and after extinction events<sup>2</sup>, or by biotic disruptions resulting from the extinctions themselves<sup>3</sup>. To the extent that extinctions entail the collapse of niches rather than the emptying of them, they restrict opportunities for future diversification<sup>16,17</sup>. Furthermore, the dynamics of evolutionary ecology are likely to be intrinsically asymmetrical; the web of ecological relationships that supports a collection of organisms may be constructed gradually, but then collapse rapidly as its component species are lost. These evolutionary mechanisms may help to explain the greater short-term inertia of origination rates as compared with diversification rates.

These results imply that evolutionary responses to extinction events are constrained by intrinsic speed limits. It has been shown previously that evolutionary responses to extinction events are delayed in time<sup>16</sup>; the current analysis shows that they are also strongly damped in amplitude. Although some groups of fossil taxa

radiate rapidly after extinction episodes<sup>1,13,18</sup>, this analysis indicates that the overall response by origination rates is attenuated, and thus biodiversity is slow to recover. If the continuing anthropogenic extinction episode<sup>19–21</sup> turns out to be comparable to those in the fossil record (which is not yet clear<sup>22,23</sup>), my analysis shows that diversification rates are unlikely to accelerate enough to keep pace with it. Thus, widespread depletion of biodiversity would probably be permanent on multimillion-year timescales. □

## Methods

### Spectral methods

Conventional Fourier methods should not be applied to the fossil record, because its stratigraphic boundaries are unevenly spaced in time. Nor should one simply even out the spacing by interpolating within each stratigraphic interval, because this introduces artefactual correlation among the interpolated points<sup>24,25</sup>. Instead, I used the date-compensated discrete Fourier transform<sup>4</sup> and the Lomb–Scargle Fourier transform<sup>5</sup>, two algorithms that have similar statistical behaviour to the conventional Fourier transform, but can be applied directly to unevenly spaced data. Both techniques yielded similar results. Each spectrum was smoothed over a window spanning about 5% of the width of the spectrum plots; this reduces the noise in the spectrum, but preserves the average spectral power at each wavelength. This is appropriate because my analysis concerns the broadband spectral properties of the fossil record rather than narrow spectral peaks, such as would be used to identify periodicities. I limited my analysis to wavelengths longer than 10 Myr (corresponding to the Nyquist frequency), because tests with synthetic non-stationary data sets show strong aliasing at shorter wavelengths.

### Source data

I used Sepkoski's compilations of fossil marine invertebrate families<sup>6</sup> and genera<sup>7</sup>, with updates through 1997. Sepkoski aggregated some stratigraphic stages and subdivided others, making his stratigraphic intervals more uniform, and thus more suitable to spectral analysis, than the conventional stratigraphic timescale. I excluded the first few intervals (> 525 Myr ago), in which total diversity is so low that the per-taxon metrics are anomalously high. I also excluded all taxa that occur in only one stratigraphic interval to minimize Lagerstätten and monographic effects<sup>7</sup>. I assigned all originations in a given interval to the stratigraphic boundary that begins it, and all extinctions to the stratigraphic boundary that ends it; other possible assignments (for example, to the midpoint of each interval) give similar results.

Received 6 June; accepted 30 October 2001.

- Erwin, D. H. The end and the beginning: recoveries from mass extinctions. *Trends Ecol. Evol.* **13**, 344–349 (1998).
- Stanley, S. M. Delayed recovery and the spacing of major extinctions. *Paleobiology* **16**, 401–414 (1990).
- Myers, N. & Knoll, A. H. The biotic crisis and the future of evolution. *Proc. Natl Acad. Sci. USA* **98**, 5389–5392 (2001).
- Foster, G. Time series analysis by projection. I. Statistical properties of Fourier analysis. *Astron. J.* **111**, 541–554 (1996).
- Scargle, J. D. Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data. *Astrophys. J.* **263**, 835–853 (1982).
- Sepkoski, J. J. Jr Ten years in the library: new data confirm paleontological patterns. *Paleobiology* **19**, 43–51 (1993).
- Sepkoski, J. J. Jr In *Global Events and Event Stratigraphy in the Phanerozoic* (ed. Walliser, O. H.) 35–51 (Springer, Berlin, 1996).
- Gilinsky, N. L. In *Analytical Paleobiology* (eds Gilinsky, N. L. & Signor, P. W.) 157–174 (The Paleontological Society, Knoxville, Tennessee, 1991).
- Foote, M. Temporal variation in extinction risk and temporal scaling of extinction metrics. *Paleobiology* **20**, 424–444 (1994).
- Newell, N. D. Revolutions in the history of life. *Geol. Soc. Am. Special Paper* **89**, 63–91 (1967).
- Smith, A. B. Large-scale heterogeneity of the fossil record: implications for Phanerozoic biodiversity studies. *Phil. Trans. R. Soc. Lond. B* **356**, 351–367 (2001).
- Foote, M. Origination and extinction components of taxonomic diversity: general problems. *Paleobiology* **26** (Suppl.), 74–102 (2000).
- Jablonski, D. Geographic variation in the molluscan recovery from the end-Cretaceous extinction. *Science* **279**, 1327–1330 (1998).
- Miller, A. I. Biotic transitions in global marine diversity. *Science* **281**, 1157–1160 (1998).
- Raup, D. M. Large-body impact and extinction in the Phanerozoic. *Paleobiology* **18**, 80–88 (1992).
- Kirchner, J. W. & Weil, A. Delayed biological recovery from extinctions throughout the fossil record. *Nature* **404**, 177–180 (2000).
- Erwin, D. H. Lessons from the past: Biotic recoveries from mass extinctions. *Proc. Natl Acad. Sci. USA* **98**, 5399–5403 (2001).
- Foote, M. Origination and extinction components of taxonomic diversity: Paleozoic and post-Paleozoic dynamics. *Paleobiology* **26**, 578–605 (2000).
- Pimm, S. L., Russell, G. J., Gittleman, G. L. & Brooks, T. M. The future of biodiversity. *Science* **269**, 347–354 (1995).
- Ehrlich, P. R. & Wilson, E. O. Biodiversity studies: science and policy. *Science* **253**, 758–762 (1991).
- Wilson, E. O. *The Diversity of Life* (Norton, New York, 1999).
- Jablonski, D. In *Extinction Rates* (eds Lawton, J. H. & May, R. M.) 25–44 (Oxford Univ. Press, New York, 1995).
- Jablonski, D. Lessons from the past: Evolutionary impacts of mass extinctions. *Proc. Natl Acad. Sci. USA* **98**, 5393–5398 (2001).

**Table 2 Log-log slopes of fossil power spectra**

| Taxonomic level and metric | Power-law slope |              |
|----------------------------|-----------------|--------------|
|                            | Extinctions     | Originations |
| Families                   |                 |              |
| per interval               | 0.04            | 0.73**       |
| per Myr                    | –0.09           | 0.89***      |
| % per interval             | 0.42*           | 1.15***      |
| % per Myr                  | 0.54*           | 1.37***      |
| Genera                     |                 |              |
| per interval               | 0.30            | 0.94***      |
| per Myr                    | 0.27            | 0.93***      |
| % per interval             | 0.35            | 0.94***      |
| % per Myr                  | 0.52*           | 1.08***      |

Statistical significance levels were estimated by repeating the same analysis on 10<sup>4</sup> randomly reshuffled copies of each data set: \*P < 0.05, \*\*P < 0.01, \*\*\*P < 0.001.

24. Schulz, M. & Stettin, K. SPECTRUM: Spectral analysis of unevenly spaced paleoclimatic time series. *Comput. Geosci.* **23**, 929–945 (1997).
25. Kirchner, J. W. & Weil, A. No fractals in fossil extinction statistics. *Nature* **395**, 337–338 (1998).
26. Cartwright, M. *Fourier Methods for Mathematicians, Scientists, and Engineers* (ed. Bell, G. M.) (Ellis Horwood, Chichester, UK, 1990).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com>).

## Acknowledgements

I am indebted to the late J. Sepkoski for his fossil databases, and to A. Weil for directing my attention to this topic and offering extensive comments on the manuscript. This work was partially supported by the National Science Foundation and the University of California.

Correspondence and requests for materials should be addressed to the author (e-mail: [kirchner@seismo.berkeley.edu](mailto:kirchner@seismo.berkeley.edu)).

## Resource-based niches provide a basis for plant species diversity and dominance in arctic tundra

Robert B. McKane\*, Loretta C. Johnson†, Gaius R. Shaver‡, Knute J. Nadelhoffer‡, Edward B. Rastetter‡, Brian Fry§, Anne E. Giblin‡, Knut Kielland||, Bonnie L. Kwiatkowski‡, James A. Laundre‡ & Georgia Murray¶

\* US Environmental Protection Agency, Corvallis, Oregon 97333, USA

† Division of Biology, Kansas State University, Manhattan, Kansas 66506, USA

‡ The Ecosystems Center, Marine Biological Laboratory, Woods Hole, Massachusetts 02543, USA

§ Institute of Pacific Islands Forestry, USDA Forest Service, Honolulu, Hawaii 96813, USA

|| Institute of Arctic Biology, University of Alaska, Fairbanks, Alaska 99775, USA

¶ Appalachian Mountain Club, Gorham, New Hampshire 03581, USA

Ecologists have long been intrigued by the ways co-occurring species divide limiting resources. Such resource partitioning, or niche differentiation, may promote species diversity by reducing competition<sup>1,2</sup>. Although resource partitioning is an important determinant of species diversity and composition in animal communities<sup>3</sup>, its importance in structuring plant communities has been difficult to resolve<sup>4</sup>. This is due mainly to difficulties in studying how plants compete for belowground resources<sup>5</sup>. Here we provide evidence from a <sup>15</sup>N-tracer field experiment showing that plant species in a nitrogen-limited, arctic tundra community were differentiated in timing, depth and chemical form of nitrogen uptake, and that species dominance was strongly correlated with uptake of the most available soil nitrogen forms. That is, the

most productive species used the most abundant nitrogen forms, and less productive species used less abundant forms. To our knowledge, this is the first documentation that the composition of a plant community is related to partitioning of differentially available forms of a single limiting resource.

Species-rich plant communities seemingly contradict the competitive exclusion principle, a fundamental tenet of ecology that predicts that species occupying the same niche cannot coexist, or that the number of species cannot exceed the number of limiting resources<sup>6,7</sup>. Plants have nothing comparable to the 'food niche' of animals. At the global level the 300,000 terrestrial plant species may have only 20 different limiting resources (light, water, CO<sub>2</sub> and the same set of mineral nutrients), and field experiments indicate that at most three or four resources are limiting in any plant community<sup>2</sup>. Consequently, some assert that resource partitioning is unimportant in maintaining species diversity in plant communities<sup>8,9</sup>.

We studied tussock tundra in arctic Alaska where plant growth is primarily limited by soil nitrogen availability<sup>10</sup>. Despite this shared limitation, species with graminoid, deciduous shrub, evergreen shrub, cryptogam and forb growth forms typically co-occur in tussock tundra at scales of less than 0.1 m<sup>2</sup> (ref. 11). This 'paradox of diversity' could be resolved if variations in resource acquisition in space and time among plant species lead to resource partitioning<sup>12</sup>, thereby allowing coexistence on a single limiting resource. There is abundant evidence that tundra plant species differ in rooting depth<sup>13</sup>, phenology<sup>14</sup>, and uptake preferences for different chemical forms of nitrogen (ammonium, nitrate and a variety of free amino acids)<sup>13,15</sup>. Similar interspecific differences have been observed for many other plant communities<sup>12,16,17</sup>. However, evidence that such differences contribute to resource partitioning within communities is limited<sup>18</sup>.

We used <sup>15</sup>N-tracers in tussock tundra at Toolik Lake, Alaska (68° 38' N, 149° 34' W, elevation 760 m), to address two questions: how are species spatially, temporally and chemically differentiated in their use of soil nitrogen and are such niche differences consistent with community structure? We hypothesized that if niche differentiation reduces competition for soil nitrogen, its signature should be evident in patterns of species dominance (productivity). That is, dominant species should use the most available forms of soil nitrogen and subordinate species less available forms.

Soluble nitrogen fractions in tundra soils vary seasonally and are strongly dominated by organic forms, including free amino acids, with lower concentrations of ammonium and nitrate<sup>19</sup>. To trace how tundra species differ in uptake of different soil nitrogen sources, we injected three chemical forms (ammonium, nitrate and glycine) of <sup>15</sup>N-labelled nitrogen at two soil depths (3 and 8 cm) twice (24 June and 7 August) in a growing season. Separate 1.5 m × 1.5 m plots were used for each of the 12 treatment combinations of chemical form, depth and time (3 × 2 × 2 factorial design). For each of the five most abundant vascular plant species (Table 1), we measured <sup>15</sup>N tracers in aboveground tissues seven days after treatment for replicate plants (*n* = 3 to 6) in each plot. Uptake of the form of available nitrogen corresponding to the <sup>15</sup>N label in each treatment

**Table 1 Plant species analysed for uptake of <sup>15</sup>N in tussock tundra at Toolik Lake, Alaska**

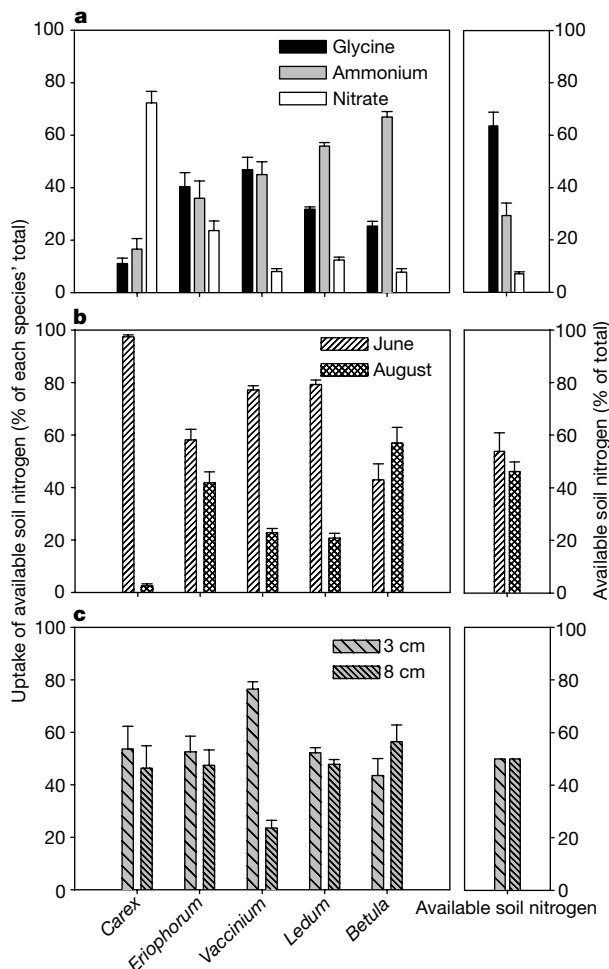
| Species*                     | Growth form     | Peak aboveground biomass<br>(% of community total)† | Aboveground net primary production<br>(% of community total)‡ |
|------------------------------|-----------------|-----------------------------------------------------|---------------------------------------------------------------|
| <i>Carex bigelowii</i>       | Graminoid       | 1 ± 1                                               | 3 ± 1                                                         |
| <i>Eriophorum vaginatum</i>  | Graminoid       | 19 ± 6                                              | 40 ± 14                                                       |
| <i>Vaccinium vitis-idaea</i> | Evergreen shrub | 18 ± 3                                              | 10 ± 2                                                        |
| <i>Ledum palustre</i>        | Evergreen shrub | 32 ± 4                                              | 16 ± 2                                                        |
| <i>Betula nana</i>           | Deciduous shrub | 19 ± 5                                              | 16 ± 4                                                        |

\* The five species described here accounted for 89% of total community aboveground biomass and 84% of net primary production. Of the 15 vascular plant species we found in this community, these were the only species that occurred in all 12 treatments. There were an average of seven vascular species per quadrat within the 20 cm × 20 cm control quadrats used to measure biomass and productivity. On average, four out of the five species in Table 1 occurred in the same quadrats. Species' root systems were closely intertwined within the organic soil layer.

† Peak (August) total aboveground vascular plant biomass = 348 ± 30 g dry weight per m<sup>2</sup>.

‡ Total aboveground vascular plant net primary production = 163 ± 21 g dry weight per m<sup>2</sup> per yr.

Data are shown ± 1 standard error.



**Figure 1** Aboveground uptake of available soil nitrogen by the five most common species in tussock tundra. Data is shown for chemical form, time and depth (left panels in **a**, **b** and **c**, respectively). The corresponding pools of available (1 M KCl-extractable) soil nitrogen are shown for comparison (right panels in **a**, **b** and **c**). Data are expressed as a percentage of each species' total uptake (left panels), or as a percentage of total available nitrogen (right panels). Data were summarized from Table 2 as the pooled means and standard errors for treatments having a common  $^{15}\text{N}$  label (chemical form, depth or time).

was calculated using an isotopic mixing model that included three variables: the amount of  $^{15}\text{N}$  label applied, the amount of the corresponding source of available nitrogen estimated to be within the diffusion zone of the  $^{15}\text{N}$  label, and the measured uptake of  $^{15}\text{N}$  label (Methods, equation (2)). To normalize, we expressed available nitrogen uptake per treatment as a percentage of each species' total uptake across all 12 treatments (Table 2).

The five most common species were well differentiated in chemical form, timing, and depth of available nitrogen uptake (left frames in Fig. 1a–c). Only *Carex bigelowii* Torr. (Bigelow sedge) used mainly nitrate. *Eriophorum vaginatum* L. (cottongrass) and *Vaccinium vitis-idaea* L. (low-bush cranberry) relied mainly on glycine and ammonium; however, *Vaccinium* obtained more of these forms earlier in the growing season and at a shallower depth than did *Eriophorum*. *Ledum palustre* L. (Labrador tea) and *Betula nana* L. (dwarf birch) used mainly ammonium but differed in timing of uptake. Thus each species occupied a different niche with respect to nitrogen use. We cannot be certain that some portion of tracer nitrogen was not microbially transformed to other compounds before plant uptake (for example, glycine-N may be transformed to ammonium-N or vice versa). Nonetheless, the data indicate that species used the original tracers in fundamentally different ways. Moreover, patterns of tracer use should also reflect how native forms of soluble nitrogen are transformed and ultimately partitioned among species.

Comparison of species' uptake patterns to the pattern of available nitrogen across chemical forms, seasons and depths (left versus right frames in Fig. 1a–c) suggests that dominance is related to a species' ability to exploit the most abundant resources. *Eriophorum*, the most productive species in this community, used nitrogen in a combination of chemical forms and seasons that closely matched the most available forms of nitrogen (glycine and ammonium). In contrast, *Carex*, the least productive of five dominants (Table 1), primarily used the least available nitrogen form (nitrate) and did not compete for nitrogen later in the growing season. More precisely, *Eriophorum* and *Carex* had patterns of uptake that were 71 and 31% similar, respectively, to the pattern of nitrogen availability across all treatments, as calculated from:

$$S_i = 100 - 0.5 \sum_{h=1}^T |u_h - a_h| \quad (1)$$

where, for species  $i$ ,  $S$  is the percentage similarity between the pattern of uptake and the pattern of available nitrogen across  $T = 12$  treatments,  $u_h$  is the percentage of species  $i$ 's total uptake of available nitrogen in treatment  $h$ , and  $a_h$  is the percentage of total available nitrogen in treatment  $h$  (Table 2)<sup>12</sup>. Across all five species, productivity increased hyperbolically with the similarity between the pattern of uptake and pattern of available nitrogen ( $r^2 = 0.96$ ,  $P = 0.004$ ; Fig. 2). We tested the sensitivity of this result to uncertainties in our estimates of available glycine and the depth distribution of all three chemical nitrogen forms (see Methods). This analysis showed the relationship between productivity and  $S$  was very robust with respect to these uncertainties (Fig. 2).

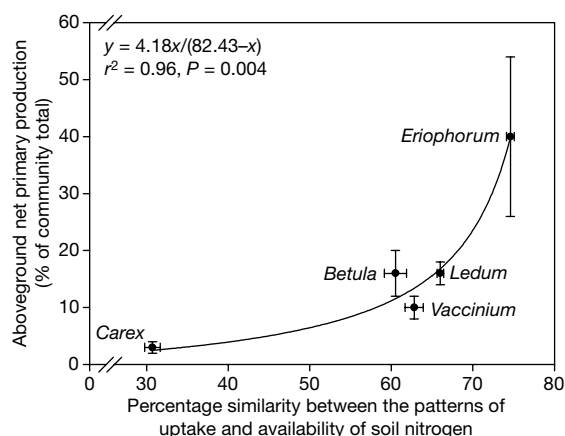
Other studies<sup>13,16–18</sup> have shown spatial, temporal or chemical differences in resource use among co-occurring plant species, but have not characterized mechanisms by which those differences

**Table 2** Species' uptake and available amounts of native soil nitrogen for  $^{15}\text{N}$ -labelled treatments

| Species            | Percentage values |      |        |      |          |      |        |      |         |      |        |      | Total across treatments<br>( $\mu\text{g m}^{-2}$ ) | I.s.d.* |
|--------------------|-------------------|------|--------|------|----------|------|--------|------|---------|------|--------|------|-----------------------------------------------------|---------|
|                    | Glycine           |      |        |      | Ammonium |      |        |      | Nitrate |      |        |      |                                                     |         |
|                    | June              |      | August |      | June     |      | August |      | June    |      | August |      |                                                     |         |
|                    | 3 cm              | 8 cm | 3 cm   | 8 cm | 3 cm     | 8 cm | 3 cm   | 8 cm | 3 cm    | 8 cm | 3 cm   | 8 cm |                                                     |         |
| Nitrogen uptake    |                   |      |        |      |          |      |        |      |         |      |        |      |                                                     |         |
| <i>Carex</i>       | 1                 | 9    | 0      | 0    | 8        | 5    | 0      | 0    | 39      | 36   | 1      | 1    | 3                                                   | 17      |
| <i>Eriophorum</i>  | 17                | 13   | 2      | 9    | 9        | 11   | 11     | 5    | 6       | 3    | 6      | 8    | 2                                                   | 10      |
| <i>Vaccinium</i>   | 35                | 5    | 6      | 3    | 25       | 6    | 7      | 6    | 3       | 3    | 1      | 0    | 13                                                  | 9       |
| <i>Ledum</i>       | 14                | 9    | 3      | 5    | 22       | 23   | 6      | 5    | 6       | 5    | 1      | 1    | 23                                                  | 5       |
| <i>Betula</i>      | 9                 | 6    | 6      | 5    | 10       | 11   | 16     | 30   | 3       | 1    | 2      | 1    | 4                                                   | 12      |
| Available nitrogen | 16                | 16   | 16     | 16   | 8        | 8    | 6      | 6    | 3       | 3    | 1      | 1    | 4,979                                               | –       |

\* Least significant difference (I.s.d.,  $P = 0.05$ ) is given for comparing treatment means within rows.





**Figure 2** Aboveground net primary production (mean  $\pm$  standard error) of the five most common species in tussock tundra. Data is shown as a function of the percentage similarity ( $S$ ) between the pattern of uptake and pattern of available soil nitrogen across three chemical forms, two seasons and two depths.  $S$  was calculated by applying equation (1) to the data in Table 2. Standard errors for  $S$  were calculated from a sensitivity analysis in which we assigned a maximum uncertainty of  $\pm 50\%$  for estimated amounts of total available glycine and the depth distribution of available glycine, ammonium and nitrate (see Methods).

influence species diversity and composition. We have proposed that if niche differentiation promotes coexistence by 'sufficiently' reducing competition, its signature should be evident in the pattern of species dominance. Our results clearly support this hypothesis by showing that species dominance in tussock tundra is positively correlated with use of the most available soil nitrogen resources.

Our results are consistent with solution culture studies showing differential uptake of ammonium and free amino acids by excised roots<sup>13</sup> or intact plants of these species in isolation<sup>15</sup>. However, because we measured nitrogen uptake in a competitive environment, our study and those conducted in solution culture are not directly comparable. In particular, uptake preferences of species isolated in solution culture should more closely reflect species' fundamental niches<sup>20</sup>. For example, when isolated in solution culture, *Ledum*, *Eriophorum*, *Betula* and *Carex* all preferred glycine over ammonium, albeit in different ratios (the glycine to ammonium uptake was 4.31, 1.45, 1.32 and 1.05, respectively)<sup>13</sup>. Our study shows that of these four species only *Eriophorum* took up more glycine than ammonium (ratio, 1.14) in the field, suggesting that our *in situ* measurements estimate realized niches<sup>20</sup>. Taken together, these studies are consistent with the idea that dominance is the appropriation of potential (fundamental) niche space of subordinate species by other dominant species<sup>21</sup>. Thus, *Eriophorum* appears to have pre-empted the most favourable niches in this community.

Although our results clearly support the niche diversification hypothesis<sup>1</sup>, our focus on the most abundant species in this community has emphasized factors (the competitive interactions involving *Eriophorum*) that are sufficiently common to have predictable effects on community structure. The importance of niche differentiation in facilitating the coexistence of the less abundant plant species in this community is an important, unresolved question. For example, the occurrences of rare species may depend primarily on infrequent or less predictable factors, such as random gaps in the spacing of dominant species, spatial heterogeneity in soil nitrogen resources, and small-scale disturbances<sup>9</sup>. However, across all 15 vascular plant species in this community, species productivity generally decreased in a rank-ordered geometric series ( $r^2 = 0.96$ ,  $P < 0.001$ ). This dominance pattern is commonly associated with plant communities of severe environments and is generally thought to be a result of niche pre-emption. In contrast, species-rich plant communities in temperate and tropical environments more commonly fit a lognormal distribution, a

dominance pattern that may reflect a larger number of controlling factors<sup>22</sup>. Our study provides a clearer view of belowground competitive interactions, and establishes a means to better examine fundamental questions about plant species diversity. □

## Methods

### Field site

This experiment was conducted in 1992 in moist tussock tundra near Toolik Lake, Alaska, at the Arctic Tundra Long Term Ecological Research (LTER) Site in the northern foothills of the Brooks Range. The soil is a histic pergelic cryaquept<sup>23</sup> and consists of a 10–50-cm peat layer over a silty mineral soil and permafrost. During the June–August growing season, mean air temperature is around 9.3 °C, total precipitation is about 180 mm, and the depth of thaw moves from near the surface to just below the organic soil layer. The vegetation at the study site<sup>11</sup> is similar to tussock tundra across the Alaskan North Slope, northern Canada, and eastern Siberia<sup>24</sup>.

### Plant biomass and productivity

Species biomass ( $\text{g m}^{-2}$ ) and net primary production ( $\text{g m}^{-2} \text{ yr}^{-1}$ ) were measured on 26–28 June and 8–13 August to coincide with the times of  $^{15}\text{N}$  labelling. At each time, total aboveground biomass was destructively measured for 16 randomly located 20 cm  $\times$  20 cm control quadrats near the  $^{15}\text{N}$  plots. Aboveground net primary production was determined from biomass of tissues produced during the current year, including leaves, apical stems and inflorescences, plus a separate measure of secondary stem production at this site<sup>25</sup>.

### $^{15}\text{N}$ treatments

The twelve  $^{15}\text{N}$  treatment plots were randomly located along a transect parallel to the slope contour (slope angle,  $\sim 5^\circ$ ). Plots were separated by at least 2 m. We injected the  $^{15}\text{N}$  (99 atom percentage  $^{15}\text{N}$ ) in solution ( $11 \text{ mmol l}^{-1}$ ) as ammonium chloride, potassium nitrate or glycine. A syringe fitted with a four-sideport needle was used to deliver 2 ml to the desired depth (3 or 8 cm) at each of 431 injection sites per plot (test injections with dye showed a diffusion radius of about 1.5 cm around points of injection). The injection sites were spaced in a 7.5 cm  $\times$  7.5 cm grid, with alternate rows offset to distribute the  $^{15}\text{N}$  more evenly. This spacing delivered 65 mg  $^{15}\text{N}$  per  $\text{m}^2$ .

### Plant $^{15}\text{N}$ analyses

Seven days after  $^{15}\text{N}$  injection, the aboveground biomass of up to six replicate plants of each of the five dominant species (Table 1) was randomly sampled from the central 1 m  $\times$  1 m area of each 1.5 m  $\times$  1.5 m treatment plot. We used a Finnigan MAT Delta S isotope mass spectrometer to determine atom percentage  $^{15}\text{N}$  on dried, milled plant samples. Background nitrogen content and natural abundance of  $^{15}\text{N}$  were determined from replicate samples ( $n = 3$ ) of each species collected from the 20 cm  $\times$  20 cm control quadrats.

### Available nitrogen

Native concentrations ( $\text{g N per g dry soil}$ ) of available ammonium, nitrate and glycine were determined for the upper 10 cm of organic soil in mid-June, mid-July and mid-August for a related study near the  $^{15}\text{N}$  plots (K.K., unpublished data). Concentrations for the 24 June and 7 August  $^{15}\text{N}$  treatment dates were linearly interpolated from the concentrations for the three times of soil sampling. Ammonium and nitrate concentrations were measured by autoanalyser methods on 1 M KCl extracts. Glycine concentrations were measured by high-performance liquid chromatography (HPLC) on water extracts, a method that does not include total exchangeable (KCl-extractable) glycine. Other work has shown that concentrations of water extractable glycine are 5 to 20 times lower than concentrations of KCl-extractable glycine (T. Nasholm, unpublished data; ref. 26). As an initial estimate, we multiplied our glycine data by 10 to express availabilities of all three forms of nitrogen on a 1 M KCl-extractable basis. Also as an initial estimate, we assumed total available nitrogen was equally distributed by depth within the 10-cm soil layer. To test the sensitivity of our results to uncertainties in these two estimates, we recalculated each species' uptake of available nitrogen (equation (2)) and  $S$  (equation (1)) for all combinations of a 0, +50 or –50% change in available glycine and the depth distribution of available glycine, ammonium and nitrate (a 50% decrease in available nitrogen for the 3-cm treatment depth corresponded to a 50% increase for the 8-cm treatment depth and vice versa).

### Nitrogen uptake calculations

Species aboveground uptake of the  $^{15}\text{N}$  labels ( $\mu\text{g m}^{-2}$ ) was calculated from data on nitrogen content, atom percentage  $^{15}\text{N}$  in excess of natural abundance, and biomass per  $\text{m}^2$ . Uptake of the form of available nitrogen corresponding to the treatment  $^{15}\text{N}$  label was calculated using an isotopic mixing model:

$$U_{\text{unlabelled}} = U_{\text{labelled}}(m_{\text{unlabelled}}/m_{\text{labelled}}) \quad (2)$$

where  $m_{\text{labelled}}$  is the total mass ( $\text{g m}^{-2}$ ) of  $^{15}\text{N}$ -labelled nitrogen injected per treatment,  $m_{\text{unlabelled}}$  is the mass of the 'target' pool of available nitrogen estimated to be within the diffusion zone of injected  $^{15}\text{N}$  (for example, the mass of 1 M KCl-extractable ammonium within a 1.5-cm spherical radius of injections of  $^{15}\text{N}$ -ammonium at the 3-cm soil depth in late June),  $U_{\text{labelled}}$  is uptake ( $\text{g m}^{-2}$ ) of  $^{15}\text{N}$  from the source  $m_{\text{labelled}}$ , and  $U_{\text{unlabelled}}$  is uptake of available nitrogen from the source  $m_{\text{unlabelled}}$ .

## Treatment comparisons

The calculated nitrogen uptake data (Table 2) were used to compare how species differ in their patterns of uptake across treatments. We emphasize that the data cannot be used to compare how species partition total nitrogen uptake within treatments because the total amount of  $^{15}\text{N}$  uptake by each species strongly depends on the relationship between lateral rooting distance and the spacing of  $^{15}\text{N}$  injections. In this experiment, where  $^{15}\text{N}$  injections were spaced in a 7.5-cm grid, *Ledum* took up six times more  $^{15}\text{N}$  across all treatments than *Eriophorum* ( $^{15}\text{N}$  data not shown). In a different experiment designed to determine species' lateral rooting distances, we injected a constant amount of  $^{15}\text{N}$  (114 mg as a mixture of 25 mmol  $\text{l}^{-1}$  each of glycine, ammonium and nitrate) at different treatment radii (5, 10, 20, 40, 60 and 100 cm) around individual plants of *Eriophorum* and *Ledum*. That design allowed  $^{15}\text{N}$  uptake data to be spatially integrated across all treatment radii, simulating how  $^{15}\text{N}$  would be taken up from injections spaced infinitely close together over an area with a radius of 100 cm. Although those results showed that the primary lateral radius of uptake (about 60% of total uptake) of *Eriophorum* was about half that of *Ledum* (5 versus 10 cm), *Eriophorum* took up almost twice as much  $^{15}\text{N}$  as *Ledum* when the data were extrapolated on a community basis (uptake per unit ground area) (R.B.M., unpublished data). The results of both experiments show that widely spaced tracer injections disproportionately label species with larger rooting areas but lower uptake capacities per unit ground area (for example, *Ledum*). In contrast, species' patterns of uptake across treatments (within rows), as presented here, should be relatively robust with respect to the spacing of tracer injections.

Received 20 September; accepted 8 November 2001.

- Hutchinson, G. E. Homage to Santa Rosalia, or why are there so many kinds of animals? *Am. Nat.* **93**, 145–159 (1959).
- Tilman, D. *Resource Competition and Community Structure* (Princeton Univ. Press, Princeton, New Jersey, 1982).
- Schoener, T. W. Resource partitioning in ecological communities. *Science* **185**, 27–39 (1974).
- Silvertown, J., Dodd, M. E., Gowing, D. J. G. & Mountford, J. O. Hydrologically defined niches reveal a basis for species richness in plant communities. *Nature* **400**, 61–63 (1999).
- Casper, B. B. & Jackson, R. B. Plant competition underground. *Ann. Rev. Ecol. Syst.* **1997**, 545–570 (1997).
- Stewart, F. M. & Levin, B. R. Partitioning of resources and the outcome of interspecific competition: a model and some general considerations. *Am. Nat.* **107**, 171–198 (1973).
- Armstrong, R. A. & McGehee, R. Competitive exclusion. *Am. Nat.* **115**, 151–170 (1980).
- Connell, J. Diversity in tropical rainforests and coral reefs. *Science* **199**, 1302–1310 (1978).
- Huston, M. A general hypothesis of species diversity. *Am. Nat.* **113**, 81–101 (1979).
- Shaver, G. R. & Chapin, F. S., III. Response to fertilization by various plant growth forms in an Alaskan tundra: nutrient accumulation and growth. *Ecology* **61**, 662–675 (1980).
- Shaver, G. R. & Chapin, F. S., III. Production: biomass relationships and element cycling in contrasting arctic vegetation types. *Ecol. Monogr.* **6**, 1–31 (1991).
- Parrish, J. A. D. & Bazzaz, F. A. Underground niche separation in successional plants. *Ecology* **57**, 1281–1288 (1976).
- Kielland, K. Amino acid absorption by arctic plants: implications for plant nutrition and nitrogen cycling. *Ecology* **75**, 2373–2383 (1994).
- Shaver, G. R. & Billings, W. D. Root production and root turnover in a wet tundra ecosystem, Barrow, Alaska. *Ecology* **56**, 401–410 (1975).
- Chapin, F. S., III, Moilanen, L. & Kielland, K. Preferential use of organic nitrogen for growth by a non-mycorrhizal arctic sedge. *Nature* **361**, 150–153 (1993).
- Flanagan, L. B., Ehleringer, J. R. & Marshall, J. D. Differential uptake of summer precipitation among co-occurring trees and shrubs in a pinyon-juniper woodland. *Plant Cell Environ.* **15**, 831–836 (1992).
- Raab, T. K., Lipson, D. A. & Monson, R. K. Soil amino acid utilization among species of the Cyperaceae: plant and soil processes. *Ecology* **80**, 2408–2419 (1999).
- McKane, R. B., Grigal, D. F. & Russelle, M. P. Spatiotemporal differences in  $^{15}\text{N}$  uptake and the organization of an old-field plant community. *Ecology* **71**, 1126–1132 (1990).
- Kielland, K. Landscape patterns of free amino acids in arctic tundra soils. *Biogeochemistry* **31**, 85–98 (1995).
- Hutchinson, G. E. Concluding remarks. *Cold Spring Harbor Symp. Quant. Biol.* **22**, 415–427 (1957).
- McNaughton, S. J. & Wolf, L. L. Dominance and the niche in ecological systems. *Science* **167**, 131–139 (1970).
- Whittaker, R. H. *Communities and Ecosystems* (Macmillan, New York, 1975).
- Rieger, S., Schoephorster, D. B. & Furbush, C. E. *Exploratory Soil Survey of Alaska* (US Department of Agriculture, Soil Conservation Service, Washington DC, 1979).
- Bliss, L. C. & Matveyeva, N. V. in *Arctic Ecosystems in a Changing Climate* (eds Chapin, F. S., III et al.) 59–89 (Academic, San Diego, California, 1992).
- Shaver, G. R. Woody stem production in Alaskan tundra shrubs. *Ecology* **67**, 660–669 (1986).
- Owen, A. G. & Jones, D. L. Competition for amino acids between wheat roots and rhizosphere microorganisms and the role of amino acids in plant N acquisition. *Soil Biol. Biochem.* **33**, 651–657 (2001).

## Acknowledgements

We thank C. Catricala, R. Brooks, J. Compton, J. Gregg, L. Gough, T. Nasholm, S. Perakis, D. Phillips and P. Rygielwicz for comments. This work was supported by the National Science Foundation and US Environmental Protection Agency.

## Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.B.M. (e-mail: mckane.bob@epa.gov).

# Polyandrous females avoid costs of inbreeding

Tom Tregenza & Nina Wedell

Ecology and Evolution Group, School of Biology, University of Leeds, Leeds LS2 9JT, UK

Why do females typically mate with more than one male? Female mating patterns have broad implications for sexual selection<sup>1,2</sup>, speciation<sup>3</sup> and conflicts of interest between the sexes<sup>4</sup>, and yet they are poorly understood. Matings inevitably have costs<sup>5</sup>, and for females, the benefits of taking more than one mate are rarely obvious. One possible explanation is that females gain benefits because they can avoid using sperm from genetically incompatible males, or invest less in the offspring of such males<sup>6,7</sup>. It has been shown that mating with more than one male can increase offspring viability<sup>8–12</sup>, but we present the first clear demonstration that this occurs because females with several mates avoid the negative effects of genetic incompatibility<sup>13</sup>. We show that in crickets, the eggs of females that mate only with siblings have decreased hatching success. However, if females mate with both a sibling and a non-sibling they avoid altogether the low egg viability associated with sibling matings. If similar effects occur in other species, inbreeding avoidance may be important in understanding the prevalence of multiple mating.

Previous studies of the highly polyandrous<sup>14</sup> (mating with more than one male)<sup>10</sup> field cricket *Gryllus bimaculatus* have shown that polyandry is associated with increased egg hatching success<sup>10</sup>. This benefit appears to stem from males having higher fertilization success when they mate with females with whom they are genetically more compatible. As yet, the source of this incompatibility is unknown. Within natural populations the negative effects of homozygosity for deleterious recessive alleles and at loci with heterozygote advantage<sup>15</sup> mean that mating with a close relative is likely to be a major source of genetic incompatibility. We suggest that females may be able to avoid this threat to their reproductive success through some mechanism that enables them to preferentially fertilize their eggs with sperm from genetically compatible males. To test this hypothesis we conducted a study in which females were allocated matings with males of known relatedness. Our prediction is that females mating with two relatives will have low offspring viability, but that polyandrous females mating with both closely related and unrelated males will have offspring viability comparable with females only mating to unrelated males.

Blocks of four sibling females were assigned to one of four treatments, all of which involved one mating with each of two different males, either with two siblings (SS), two non-siblings (NN) or a sibling and a non-sibling in either order (SN and NS) (see Methods). After mating, the hatching success of eggs was recorded. Randomized block analysis of variance reveals a significant effect of mating treatment on proportional egg hatching success ( $F_{3,75} = 6.01$ ,  $P = 0.001$ ) (Fig. 1). Hatching success does not differ significantly between experimental blocks ( $F_{25,75} = 1.48$ ,  $P = 0.10$ ). Post hoc analysis (Tukey test) of egg hatching indicates that the significant effect of treatment is due to the lower hatching success of females mated to two siblings relative to females mated to at least one non-sibling (Fig. 1) (Tukey test: NN versus SN or NS, minimum  $P = 0.68$ ; SS mating versus other treatments, maximum  $P = 0.05$ ). This suggests that females mating with a sibling and a non-sibling have egg viability similar to that of completely outbreeding females, rather than halfway between completely outbreeding and completely inbreeding females. We can test this explicitly by comparing the hatching success of females within each block using a paired *t*-test of mean hatching success of NN

and SS females versus SN and NS females. This shows there is indeed a significant improvement in egg viability in polyandrous females mating to both related and unrelated males, relative to that which would be expected if sperm from both males were used equally ( $t = 2.53$ , degrees of freedom, d.f. = 25,  $P = 0.018$ ).

There is no significant difference in hatching success according to the order in which females mated to a sibling and a non-sibling (NS versus SN, Tukey test  $P = 0.68$ ), which confirms the lack of any effect of mating order on sperm precedence<sup>16</sup>. The overall mean hatching success across females was 48% (standard error of the mean, s.e.m. = 2.2%). This is similar to the 46% hatching success observed in females from wild populations which were caught as adults (presumably having already mated), and allowed to mate repeatedly with a single male<sup>17</sup>. It is also similar to the 47% hatching success we observed in females mating twice to each of two males in a previous study<sup>10</sup>.

There is no effect of mating treatment on the number of eggs laid in the three days following mating (randomized block analysis of variance with female body size as a covariate,  $F_{3,74} = 1.25$ ,  $P = 0.30$ ). There is a difference between blocks ( $F_{25,74} = 2.58$ ,  $P = 0.001$ ) which may be due to variation between families in their egg-laying rate (because each family was used in only one or two blocks) although any such differences cannot be separated from possible differences arising from the fact that blocks were carried out at different times. The lack of effect of mating treatment indicates that females do not refrain from oviposition when mated to siblings, a behaviour observed in two species of *Drosophila*<sup>18,19</sup>.

The finding that females mated to a brother and a non-sibling have higher egg viability than would be expected if related and unrelated males had equal fertilization success provides the first direct evidence, to our knowledge, that polyandrous females increase offspring fitness through avoidance of the negative effects of genetic incompatibility. The only plausible way this effect could arise is if there is differential fertilization success of sperm<sup>2</sup> in favour of the ejaculate of unrelated males. The alternative explanation that females mated to an unrelated male increase egg viability by allocating more resources to eggs (increasing their viability), is unconvincing. There is limited scope for differential allocation in this species since the eggs laid in the period we observed are typically already chorionated in four-day-old females and cannot be further provisioned<sup>20</sup>. Additionally, if there are other costs of inbreeding that mean females should avoid investing in eggs fertilized by siblings it would be very much more efficient for them simply to oviposit less when mated to incompatible males. Females of this

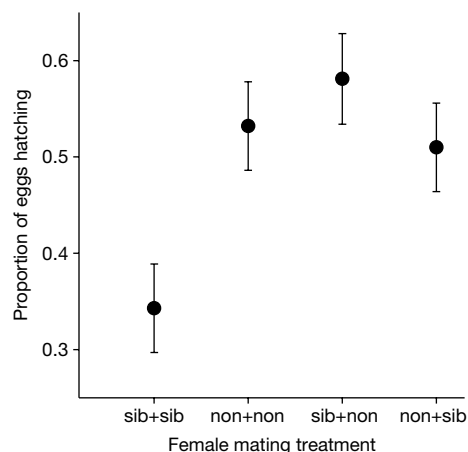
species have been shown to increase oviposition when given the opportunity to choose a mate rather than having one allocated to them<sup>21</sup>, so they are clearly capable of such behaviour.

The mechanism by which differential fertilization occurs is unknown. One possibility is that males choose to inseminate less sperm into related females. However, this is unlikely because males in our experiment had produced a spermatophore before contact with the female and hence could not manipulate the size of their ejaculate, suggesting that differential male-fertilization success is a female effect. Females from laboratory populations are able to recognize kin using olfaction<sup>22</sup>, and have been found to be less willing to mate with full siblings<sup>23</sup>. No precopulatory differences in female behaviour in relation to male relatedness were observed in our study, but females may exercise choice by accepting less sperm from closely related males, or through post-copulatory mechanisms that reduce the fertilization success of sperm from related males. Higher sperm competition success of unrelated males has been described in another species of cricket<sup>24</sup>, although the effect was not statistically significant. Analogous differences in postcopulatory fertilization success are known in matings between species or geographic races<sup>25</sup>, where sperm from more genetically distant males is frequently less successful. This is clearly a different process to that observed in this study, where more genetically similar individuals are less successful, but the existence of conspecific sperm precedence illustrates the potential for female sperm choice.

Previous studies have found evidence that mating with more than one male is associated with higher egg or offspring viability<sup>26</sup>. Field studies of adders<sup>8</sup> and sand lizards<sup>9</sup> have shown correlations between female promiscuity and offspring viability. Experiments using pseudoscorpions<sup>11</sup>, field crickets<sup>10,27</sup> and cuis<sup>12</sup>, controlling number of matings and allocating pairings at random to rule out precopulatory choice, have also found viability benefits of polyandry. Because these studies fail to find any evidence for 'viability genes' they provide indirect support for benefits of polyandry due to avoidance of genetic incompatibility. However, none of them attempted to manipulate genetic compatibility and therefore direct evidence for the hypothesis is lacking. The best existing evidence that within-population polyandry may allow females to avoid genetic incompatibility comes from the sand lizard<sup>28</sup>, in which there is a negative correlation between male relatedness and success in sperm competition. Although no direct link between differential success in sperm competition and increased offspring fitness has been demonstrated in sand lizards, the existence of inbreeding depression<sup>29</sup> suggests it may occur.

A previous study of the cricket *Teleogryllus oceanicus*<sup>27</sup> attempted to test the genetic incompatibility hypothesis through the prediction of a positive association between hatching success and paternity skew. No such relationship was found in 16 females, each mated to 2 males. However, the prediction that females biasing offspring paternity to a greater extent will have higher offspring viability is much more difficult to test if there is variation in the relative compatibility of mates. If some females are mated to two compatible males they will have high hatching success even if they do not bias paternity, whereas a female mated to two incompatible males could bias paternity completely in favour of the slightly more compatible of the two but still have low hatching success. Male compatibility was not manipulated in this previous study<sup>27</sup>, so it is difficult to exclude genetic incompatibility as an explanation for the observed viability benefits of polyandry.

It is not known how common matings between full siblings are in natural populations of crickets, although the large numbers of eggs produced by a single female certainly create the potential for sibling matings. The substantial fitness benefits of polyandry in females exposed to brothers and unrelated males suggests that selection could favour polyandry even if sibling matings are rare, or if polyandrous females can avoid the lesser, but still appreciable costs of inbreeding with more distantly related males. If the ability



**Figure 1** Relative hatching success. Eggs were from female field crickets mated either to two siblings, two non-siblings, a sibling followed by a non-sibling or a non-sibling followed by a sibling (means and standard errors).



of female field crickets to avoid using sperm from related males is shared by other species, this form of genetic incompatibility avoidance may be an important factor promoting female promiscuity across taxa. □

## Methods

### Crickets

All individuals were  $F_2$  descendants of gravid female crickets collected from the wild in Gabarone, Botswana. Offspring of the parental females were kept separately and one virgin  $F_1$  female from each parental female was mated to a single male from another family to create a set of unrelated full-sibling families. It is possible that females in the original collection may include relatives; this would be conservative in relation to our study. Families were reared in 131 plastic cages at 29°C and 18:6 hours light:dark and freely provided with rodent food pellets and water. Experimental individuals were collected from families as late-instar nymphs to ensure virginity and isolated in separate 9-cm-diameter pots provided with food and water. The experiment was arranged in blocks of four females and four males; six of these were siblings (four females and two males from one family), and two were male siblings from another family.

### Mating

All females were virgins; all males had mated once to an unrelated female on the previous day. Only males who had a spermatophore ready for transfer were used: such males produce courtship song as soon as they contact a female<sup>30</sup>. Males were used in one block only. Matings were allowed by adding the male to a 9-cm-diameter pot containing a female. Nearly all pairs mated within 10 minutes of being introduced. If the female did not mate within an hour of being placed with her first mate she was replaced with a sister (5 out of 114 females). After mating the male was allowed to stay with the female for a further 45 min to prevent her from removing the spermatophore; no female removed a spermatophore before this time. An hour after her first mating the female was mated a second time to a different male using the same protocol. In each block, females were allocated to one of four treatments: (1) one mating to each of two of her brothers (SS); (2) one mating to each of two males which were brothers to one another but unrelated to the female (NN); (3) one mating to a sibling male followed by one mating to an unrelated male (SN); (4) one mating to an unrelated male followed by one mating to a sibling male (NS). Therefore, all females mated twice and all males mated twice during the experiment and once previously. Twenty-eight blocks were carried out, with 15 families each used in a maximum of two blocks. In two blocks a female died before laying eggs. These blocks were excluded from the analysis.

### Eggs

After mating, the female was placed in a 9-cm-diameter pot and provided with food and fine wet sand for oviposition. Sand was kept moist at all times, and after three days was sieved to remove eggs. Eggs were counted and placed on a wet cotton wool pad in a petri dish and maintained under the same conditions as the adults. Eggs were checked daily for hatching until seven days after the last emergence, by which time eggs that have not hatched have begun to break down. To check for changes in hatching success over time, females laying less than 100 eggs were given a further three days for oviposition, repeated for up to 12 days or until the female had laid 100 eggs. Paired  $t$ -tests of the proportional hatching of eggs laid in the first three days versus those laid in the second three days, or versus all subsequent eggs, indicated no change in hatching success over time ( $t = 0.4$ ,  $P = 0.6$ , d.f. = 31, in both cases). The cube of the length of the hind femur of all females was used as a body size measure, it did not affect egg hatching success ( $F_{1,102} = 0.005$ ,  $P = 0.94$ ), but did have an effect on number of eggs laid ( $F_{1,102} = 6.67$ ,  $P = 0.01$ ), and is included in the analysis of eggs laid as a covariate. In all analyses, proportions were arcsine transformed and numbers of eggs laid were square-root transformed to normalize their distribution. All  $P$  values are two-tailed.

Received 31 August; accepted 30 October 2001.

- Andersson, M. *Sexual Selection* (Princeton Univ. Press, Princeton, New Jersey, 1994).
- Eberhard, W. G. *Female Control: Sexual Selection by Cryptic Female Choice* (Princeton Univ. Press, Princeton, New Jersey, 1996).
- Arnqvist, G., Edvardsson, M., Friberg, U. & Nilsson, T. Sexual conflict promotes speciation in insects. *Proc. Natl Acad. Sci. USA* **97**, 10460–10464 (2000).
- Parker, G. A. in *Sexual Selection and Reproductive Competition in Insects* (eds Blum, M. S. & Blum, A. N.) 123–166 (Academic, London, 1979).
- Daly, M. The cost of mating. *Am. Nat.* **112**, 771–774 (1978).
- Zeh, J. A. & Zeh, D. W. The evolution of polyandry I: Intra-genomic conflict and genetic incompatibility. *Proc. R. Soc. Lond. B* **263**, 1711–1717 (1996).
- Zeh, J. A. & Zeh, D. W. The evolution of polyandry 2: Post-copulatory defences against genetic incompatibility. *Proc. R. Soc. Lond. B* **264**, 69–75 (1997).
- Madsen, T., Shine, R., Loman, J. & Håkansson, T. Why do female adders copulate so frequently? *Nature* **355**, 440–441 (1992).
- Olsson, M., Gullberg, A., Tegelström, H., Madsen, T. & Shine, R. Can female adders multiply? *Nature* **369**, 528 (1994).
- Tregenza, T. & Wedell, N. Benefits of multiple mates in the cricket *Gryllus bimaculatus*. *Evolution* **52**, 1726–1730 (1998).
- Newcomer, S. D., Zeh, J. A. & Zeh, D. W. Genetic benefits enhance the reproductive success of polyandrous females. *Proc. Natl Acad. Sci. USA* **96**, 10236–10241 (1999).
- Keil, A. & Sachser, N. Reproductive benefits from female promiscuous mating in a small mammal. *Ethology* **104**, 897–903 (1998).

- Tregenza, T. & Wedell, N. Genetic compatibility, mate choice and patterns of parentage. *Mol. Ecol.* **9**, 1013–1027 (2000).
- Simmons, L. W. Female choice in the field cricket *Gryllus bimaculatus* (de Geer). *Anim. Behav.* **34**, 1463–1470 (1986).
- Thornhill, N. W. *Natural History of Inbreeding and Outbreeding: Theoretical and Empirical Perspectives* (Univ. Chicago Press, Chicago, 1993).
- Simmons, L. W. Sperm competition as a mechanism of female choice in the field cricket, *Gryllus bimaculatus*. *Behav. Ecol. Sociobiol.* **21**, 197–202 (1987).
- Qayyum, H. A. & Wahla, M. M. Biology of *Gryllus bimaculatus* Degeer. *Pak. J. Sci. Res.* **21**, 112–116 (1969).
- Markow, T. A. in *Ecological Genetics and Evolution: The Cactus–Yeast–Drosophila Model System* (eds Barker, J. S. F. & Starmer, W. T.) 273–287 (Academic, New York, 1982).
- Markow, T. A. Assortative fertilization in *Drosophila*. *Proc. Natl Acad. Sci. USA* **94**, 7756–7760 (1997).
- Lorenz, M. W., Lorenz, J. L., Treiblmaier, K. & Hoffmann, K. H. In vivo effects of allatostatin in crickets, *Gryllus bimaculatus* (Ensifera: Gryllidae). *Archives Insect Biochem. Physiol.* **38**, 32–43 (1998).
- Simmons, L. W. Female choice contributes to offspring fitness in the field cricket *Gryllus bimaculatus* (de Geer). *Behav. Ecol. Sociobiol.* **21**, 313–321 (1987).
- Simmons, L. W. Pheromonal cues for the recognition of kin by female field crickets, *Gryllus bimaculatus*. *Anim. Behav.* **40**, 192–195 (1990).
- Simmons, L. W. Kin recognition and its influence on mating preferences of the field cricket, *Gryllus bimaculatus* (de Geer). *Anim. Behav.* **38**, 68–77 (1989).
- Stockley, P. Sperm selection and genetic incompatibility: does relatedness of mates affect male success in sperm competition? *Proc. R. Soc. Lond. B* **266**, 1663–1669 (1999).
- Howard, D. J. Conspecific sperm and pollen precedence and speciation. *Annu. Rev. Ecol. Syst.* **30**, 109–132 (1999).
- Jennions, M. D. & Petrie, M. Why do females mate multiply? A review of the genetic benefits. *Biol. Rev. Cambridge Phil. Soc.* **75**, 21–64 (2000).
- Simmons, L. W. The evolution of polyandry: an examination of the genetic incompatibility and good-sperm hypotheses. *J. Evol. Biol.* **14**, 585–594 (2001).
- Olsson, M., Shine, R., Madsen, T., Gullberg, A. & Tegelström, H. Sperm selection by females. *Nature* **383**, 585 (1996).
- Olsson, M., Gullberg, A. & Tegelström, H. Malformed offspring, sibling matings, and selection against inbreeding in the sand lizard (*Lacerta agilis*). *J. Evol. Biol.* **9**, 229–242 (1996).
- Tregenza, T. & Wedell, N. Definitive evidence for cuticular pheromones in a cricket. *Anim. Behav.* **54**, 979–984 (1997).

## Acknowledgements

We thank B. Riddoch for collecting the original field population. A. Bretman and F. A. Attia for help with experiments, J. Kotiaho for suggestions on experimental design and G. Arnqvist, R. K. Butlin, D. Hosken, L. Rowe, R. Wilson and D. Zeh for comments on the manuscript. This work was supported by a NERC fellowship to T.T. and a Royal Society fellowship to N.W.

## Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.T. (e-mail: T.Tregenza@leeds.ac.uk).

# The optic tectum controls visually guided adaptive plasticity in the owl's auditory space map

Peter S. Hyde\*† & Eric I. Knudsen\*

\* Department of Neurobiology, Stanford University School of Medicine, Stanford, California 94305, USA

The midbrain contains an auditory map of space that is shaped by visual experience<sup>1–3</sup>. When barn owls are raised wearing spectacles that horizontally displace the visual field, the auditory space map in the external nucleus of the inferior colliculus (ICX) shifts according to the optical displacement of the prisms<sup>4</sup>. Topographic visual activity in the optic tectum could serve as the template that instructs the auditory space map<sup>5</sup>. We studied the effects of a restricted, unilateral lesion in the portion of the optic tectum that

† Present address: Science Department, Roxbury Latin School, West Roxbury, Massachusetts 02132, USA.

represents frontal space. Here we show that such a lesion eliminates adaptive adjustments specifically in the portion of the auditory map that represents frontal space on the same side of the brain, while the rest of the map continues to adjust adaptively. Thus, activity in the tectum calibrates the auditory space map in a location-specific manner. Because the site of adaptive changes is the ICX<sup>4,6,7</sup>, the results also indicate that the tectum provides a topographic instructive signal that controls adaptive auditory plasticity in the ICX.

The auditory system creates a map of space in the midbrain that helps to orient the eyes towards interesting sounds<sup>8</sup>. The map is located in the optic tectum (the superior colliculus in mammals) where it aligns with a map of visual space<sup>9–11</sup>. The alignment of these maps represents the associations of auditory localization cues, such

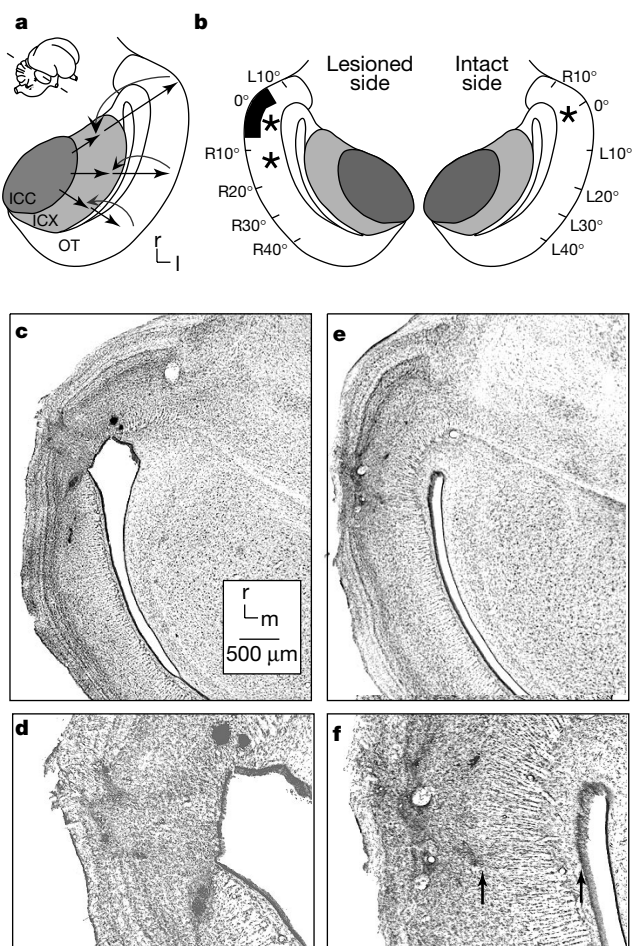
as interaural time difference (ITD), with locations in the visual field. These associations are shaped by experience<sup>12,13</sup>.

In the barn owl, the auditory space map that appears in the optic tectum is created and calibrated in the ICX<sup>4</sup>. The calibrated auditory space map is relayed to the deep layers of the optic tectum<sup>14</sup>, where it merges with the visual space map to create a multimodal representation of space<sup>9</sup>.

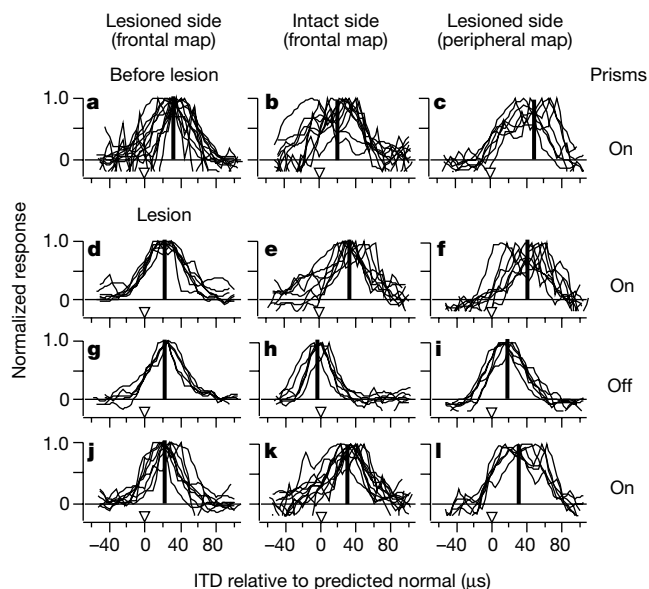
A potential source of the instructive signal that calibrates the auditory space map in the ICX is the optic tectum<sup>15</sup>. Neurons located primarily in the intermediate layers of the optic tectum send a topographic feedback projection to the site of plasticity in the ipsilateral ICX<sup>15,16</sup> (Fig. 1a). We investigated whether a lesion in the optic tectum can eliminate the adaptive auditory plasticity that normally occurs in the ICX.

In most of our experiments, we monitored the plasticity of the auditory space map in the deep layers of the optic tectum (Fig. 1). The powerful feedforward connections from the ICX to the deep tectum<sup>6</sup> cause auditory tuning in the deep tectum to follow faithfully changes that occur in the ICX<sup>4,17</sup>. There are two important reasons for monitoring the plasticity of the space map in the tectum. First, by recording at the output of the ICX, tissue damage caused by repeated electrode penetrations could not be responsible for a loss of map plasticity in the ICX. Second, sites in the deep tectum have visual receptive fields that enable precise assessment of changes in ITD tuning (Methods). Because activity in the deep tectum represents the output of the ICX, a loss of adaptive change observed in the deep tectum must correspond to a loss of plasticity in the ICX.

In each of five owls, a lesion was made on one side of the brain in the part of the map that represents frontal space. The lesions were made in the superficial tectal layers and damage extended into the deep layers (Methods). In no case, however, did a lesion encroach on the ICX (Fig. 1b–d, Supplementary Information). We monitored the effect of the lesion on auditory plasticity at three locations in the deep tectum (Fig. 1b, asterisks). Two of the locations were on the



**Figure 1** Schematics and representative lesions. **a**, Cartoon of the flow of auditory spatial information from the central nucleus of the inferior colliculus (ICC) to the external nucleus of the inferior colliculus (ICX) to the optic tectum (OT, straight arrows) and the projection from the OT to the ICX (curved arrows). The map of space represented in each tectum is indicated in degrees of azimuth relative to the owl's midline: R, right; L, left. **b**, Schematic locations of the tectal lesion (black shading) and recording sites (asterisks). **c**, Section through a lesion (owl DOB) typical of those that blocked plasticity in the ipsilateral, frontal map. **d**, High-power view of the lesion shown in **c**. We note the atrophy of the tectum and the disruption of the layers at the site of the lesion. Small dark spots in the deep layers are associated with marking lesions made during the final experiment. **e**, Section through the only lesion that did not block plasticity. **f**, High-power view of the lesion shown in **e**. Small dark spots in the deep layers are associated with marking lesions made during the final experiment. We note that the tectum is less atrophied and the deep layers, 13–15 (indicated by arrows), are clearly discernible. See also Supplementary Information. r, rostral; l, lateral; m, medial.



**Figure 2** Representative lesion effects on auditory plasticity. Interaural time difference (ITD) tuning curves of all sites measured in the frontal map (visual azimuths, without prisms, from L5° to R5°) on the lesioned side (left), the frontal map on the intact control side (middle) and the peripheral map (visual azimuths, without prisms, from 10° to 25°, contralateral) on the lesioned side (right) of owl DOB. Recordings were made before the lesion (**a–c**), 2 weeks after the lesion with prisms still in place (**d–f**), 3 weeks after prism removal (**g–i**), and 8 weeks after prism replacement (**j–l**). Vertical bars indicate the mean best ITD relative to normal (open triangles).

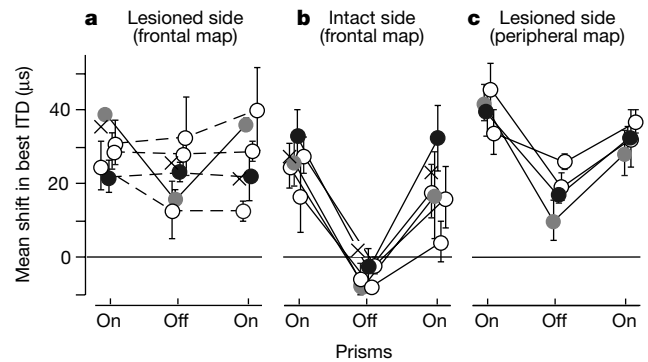
lesioned side of the brain: in the frontal map and in a more peripheral portion of the map (Fig. 2 legend). The third location was in the frontal map on the side of the brain that remained intact (Fig. 1b). The frontal region of space is represented on both sides of the brain and the connections between the ICX and optic tectum do not cross to the opposite side<sup>6,15</sup>, so we could assess plasticity in the frontal map with tectal feedback either eliminated (on the lesioned side of the brain) or intact (on the intact side). The frontal map on the intact side of the brain served as a control for variations in experience across space and across individuals.

Results from one owl are shown in Fig. 2. First, the capacity for auditory plasticity was verified by exposing the owl to prisms for six weeks before the lesion was placed. Once the capacity to shift ITD tuning had been assessed (Fig. 2a–c, Table 1, owl DOB), a lesion in the optic tectum was made (Fig. 1c, d) and the prisms were left in place. Two weeks later, with the prisms still in place, ITD tuning had shifted back towards normal in the frontal map on the lesioned side (compare Fig. 2a and d;  $P < 0.05$ , two-tailed  $t$ -test). This effect was observed in four of the five owls (Table 1). There was no such consistent effect on ITD tuning at either of the other two map locations (Table 1). This suggests that a signal in the tectum is necessary to maintain fully shifted ITD tuning, counteracting non-visual influences<sup>18</sup>, and innate preferences for normal tuning<sup>19,20</sup> that tend to drive ITD tuning in the map towards normal. This result also indicates that plasticity itself was not eliminated by the lesion.

At this point, the prisms were removed. After three weeks of normal vision (Fig. 2g–i), ITD tuning in the frontal map on the lesioned side had not changed ( $P > 0.05$ , two-tailed  $t$ -test). In contrast, the mean best ITD in the frontal map on the intact side had shifted back to normal ( $P < 0.01$ , two-tailed  $t$ -test). In the peripheral map on the lesioned side of the brain, ITD tuning had also shifted towards normal ( $P < 0.01$ , two-tailed  $t$ -test, Fig. 2i), but the shift was incomplete. The shift may have been incomplete because the lesion may have damaged visual input to the more peripheral portion of the map, although visual responses in this region remained strong. Alternatively, the shift may have been incomplete because ITD tuning in the frontal map on this side of the brain (Fig. 2g) remained fixed on the values that the peripheral map should have come to represent, hindering the acquisition of responses to these values through lateral inhibition<sup>21</sup>.

The prisms were then put back on the owl. After more than 8 weeks of prism experience (Fig. 2j–l), ITD tuning had not changed ( $P > 0.05$ , two-tailed  $t$ -test) in the frontal map on the lesioned side. In contrast, in the frontal map on the intact control side, the mean best ITD had shifted  $35 \pm 9 \mu\text{s}$  in the adaptive direction ( $P < 0.01$ , two-tailed  $t$ -test). Also, in the peripheral map on the lesioned side, the mean best ITD had shifted adaptively by  $15 \pm 6 \mu\text{s}$  ( $P < 0.01$ , two-tailed  $t$ -test), returning, approximately, to the initially shifted value.

The data from all five owls are summarized in Fig. 3. In all cases but one, we observed similar results. In the one case in which plasticity was not eliminated by the lesion, the lesion did little damage to the layers below layer 12 (Fig. 1f, Supplementary



**Figure 3** Summary of the effects of tectal lesion on auditory map plasticity. Mean best ITD shift in the frontal map on the lesioned (a, 5 owls) and intact (b, 5 owls) sides and the peripheral map on the lesioned side (c, 4 owls) with the prisms on, off and on again. Solid lines signify significant changes ( $P < 0.05$ , two-tailed  $t$ -test); dashed lines signify no change ( $P > 0.05$ , two-tailed  $t$ -test). Open circles indicate data from owls in which plasticity was eliminated in the frontal map on the lesioned side, grey circles represent data from the owl in which plasticity was not eliminated by the lesion. Black circles indicate data shown in Fig. 2d–l. Crosses indicate data from chronically recorded sites in the frontal ICX on the lesioned side (a) and frontal tectum on the intact side (b) of one owl. In this owl, the gradual decline in the shift in best ITD in the ICX following prism removal was not correlated with the change in visual experience and is probably the same drift towards normal that was observed in the lesioned, frontal tectum of other owls before the removal of the prisms (Table 1).

Information). These deep layers contain some of the neurons that project back to the ICX<sup>15,16</sup> and they receive visual and other potentially instructive input from the forebrain<sup>22</sup> that may have enabled adaptive plasticity to occur.

To test the inference that the tectal lesions eliminated plasticity in the corresponding portion of the ICX, we performed one additional experiment. In one owl that underwent the same manipulations as the other owls, we recorded chronically from a single site in the frontal ICX on the lesioned side of the brain (Methods). The results were the same: adaptive adjustments were prevented at the ICX site on the lesioned side but continued normally at a corresponding tectal site on the intact side (Fig. 3, crosses).

Precise, topographic visual activity is strong in the tectal layers where the feedback projection originates<sup>9</sup>, so the tectal signal to the ICX may simply be a topographic visual template that instructs changes by strengthening active auditory inputs at the site of visual activation and weakening active auditory inputs at other sites. Visually driven activity has not been observed in the ICX, however, perhaps because the signal from the optic tectum is gated in some way (by attention, for example).

Our results suggest that instructive activity from the visual system is sent into the auditory system to guide the transformation of localization cues into a retinotopic map of space<sup>9</sup>. The same signal is

**Table 1** Effect of lesion with prisms in place

|                                | Shift in best ITD ( $\mu\text{s}$ ) |                              |                            |                             |                             |                             |                             |                             |                            |                              |
|--------------------------------|-------------------------------------|------------------------------|----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|----------------------------|------------------------------|
|                                | Owl DOB                             |                              | Owl HED                    |                             | Owl GRY                     |                             | Owl SCA                     |                             | Owl TGA                    |                              |
|                                | Initial shift                       | After lesion                 | Initial shift              | After lesion                | Initial shift               | After lesion                | Initial shift               | After lesion                | Initial shift              | After lesion                 |
| Frontal map (lesioned side)    | $32 \pm 4$<br>( $n = 7$ )           | $23 \pm 3^*$<br>( $n = 5$ )  | $47 \pm 6$<br>( $n = 7$ )  | $38 \pm 7^*$<br>( $n = 3$ ) | $35 \pm 2$<br>( $n = 4$ )   | $29 \pm 5^*$<br>( $n = 5$ ) | $48 \pm 5$<br>( $n = 8$ )   | $31 \pm 6^*$<br>( $n = 4$ ) | $28 \pm 4$<br>( $n = 12$ ) | $24 \pm 7$<br>( $n = 3$ )    |
| Frontal map (intact side)      | $20 \pm 14$<br>( $n = 10$ )         | $33 \pm 4^*$<br>( $n = 8$ )  | $27 \pm 17$<br>( $n = 8$ ) | $26 \pm 7$<br>( $n = 6$ )   | $24 \pm 12$<br>( $n = 8$ )  | $17 \pm 10$<br>( $n = 5$ )  | $33 \pm 11$<br>( $n = 11$ ) | $27 \pm 5$<br>( $n = 6$ )   | $33 \pm 9$<br>( $n = 11$ ) | $25 \pm 6^*$<br>( $n = 10$ ) |
| Peripheral map (lesioned side) | $48 \pm 15$<br>( $n = 7$ )          | $39 \pm 12^*$<br>( $n = 8$ ) | $47 \pm 10$<br>( $n = 3$ ) | $42 \pm 5$<br>( $n = 4$ )   | $41 \pm 10$<br>( $n = 10$ ) | $46 \pm 7$<br>( $n = 7$ )   | $47 \pm 8$<br>( $n = 9$ )   | $34 \pm 6^*$<br>( $n = 5$ ) | —                          | —                            |

\* Shift in best ITD after lesion differs from initial shift in best ITD (two-tailed  $t$ -test,  $P < 0.05$ ).



presumably responsible for the adaptive adjustment of the auditory space map in response to altered hearing<sup>23–26</sup>. This instructive signal creates a common representation of space from auditory and visual information that is initially encoded in different coordinate frames. An analogous strategy could be used to guide transformations of information that occur in a wide variety of sensory and motor systems<sup>27–30</sup>. By studying how the visual system exerts its influence on the auditory localization pathway, we may learn some of the mechanisms by which the nervous system instructs such transformations. □

## Methods

### Animals

Six barn owls (*Tyto alba*) were raised in brooding boxes with their siblings until they were approximately 60 days old. Then they were surgically prepared for experiments, as described previously<sup>4</sup>, and moved to group flight rooms for the duration of the experiment.

### Visual experience

Map plasticity was induced by fitting the owls with prismatic spectacles that horizontally displaced the visual field by 17° or 23°. Displacement was maintained for 5 to 7 weeks before the initial assessment of the auditory space map. This experience leads to an adaptive shift in ITD tuning in the ICX<sup>4</sup>.

One to four weeks following the completion of the tectal lesion, the prisms were removed. Normal vision was maintained until the auditory space map returned to normal on the intact side of the brain: approximately 3 weeks. The prisms were then reattached until the auditory space map readapted to the prisms on the intact side of the brain (a period ranging from 6 to 11 weeks).

### Electrophysiology

Acute and chronic electrophysiological techniques were the same as in previous studies<sup>1–4</sup>. In the acute experiments (five owls), ITD tuning was monitored in the deep layers of the optic tectum. In the chronic experiment (one owl), ITD tuning was monitored at one site on the 0-μs transect<sup>4</sup> in the rostral ICX on the lesioned side of the brain and at one site in the rostral optic tectum on the intact side. The location of the recording site in the ICX was verified by lesion reconstruction.

Visual receptive fields were measured with the prisms removed by projecting positive or negative contrast bars onto a tangent screen in front of the owl. Because the retinal input to the tectum terminates in the superficial layers, visual (but not auditory) responses frequently became weak or unmeasurable beneath the lesion. However, because electrode penetrations were made along iso-azimuth contours in the tectum, receptive field locations could be reliably inferred from the receptive fields measured in the intact tissue sampled at the top and/or bottom of each penetration<sup>9</sup>. For each site in the optic tectum, the predicted normal best ITD was calculated using the normal relationship between visual receptive field azimuth and best ITD<sup>4</sup>.

Auditory tuning was measured by presenting sounds dichotically as described previously<sup>1–4</sup>. Tuning for ITD was assessed by presenting 15–20 series of noise bursts with ITD varied in a random, interleaved fashion while interaural level difference was held at the optimal value for the site. Responses were defined as the number of spikes recorded in the 100 ms following stimulus presentation minus the number of spikes recorded in the 100 ms before stimulus presentation. The best ITD was the midpoint of the range of ITDs over which the response exceeded 50% of the maximum response.

ITD shift in each portion of the map was assessed by recording acutely at 3–10 sites. A difference between the measured and predicted normal best ITD at a given site indicated a shift in best ITD for that site. Positive shifts signify changes in the direction predicted by the optical displacement of the prisms. The effects of each manipulation were determined in each portion of the map in each owl by comparing the mean shift in best ITD measured before the manipulation with the mean shift measured after the manipulation, using a two-tailed *t*-test.

### Lesions

The rostral optic tectum on one side of the brain was lesioned electrolytically after ITD tuning had shifted adaptively. The lesion was made in 2–3 sessions, over 1–2 weeks, in the right optic tectum of one owl wearing right-displacing prisms, and in the left optic tectum of five owls wearing left-displacing prisms. In each session, we made a grid of 20–30 penetrations, separated by 250–500 μm, using a tungsten microelectrode (0.1–2 MΩ at 1 kHz). We passed 50 μA of cathodal current for 45 s at all sites (total is 15–25) with visual receptive fields in the targeted region of space (from L5° to R5° in azimuth and 5° up to 15° down in elevation) and with bursting responses characteristic of the superficial tectal layers. The lesion was considered to be complete when bursting activity had been eliminated from the targeted portion of the tectum. Owls with tectal lesions appeared to fly, feed and interact normally.

### Histology

At the end of the experiment, marking lesions were made, the owls were killed and perfused and the brains were prepared for histology as described previously<sup>4</sup>.

Received 23 July; accepted 29 October 2001.

- Knudsen, E. I. & Brainard, M. S. Visual instruction of the neural map of auditory space in the developing optic tectum. *Science* **253**, 85–87 (1991).
- King, A. J. A map of auditory space in the mammalian brain: neural computation and development. *Exp. Physiol.* **78**, 559–590 (1993).
- Withington-Wray, D. J., Binns, K. E. & Keating, M. J. The maturation of the superior collicular map of auditory space in the guinea pig is disrupted by developmental visual deprivation. *Eur. J. Neurosci.* **2**, 682–692 (1990).
- Brainard, M. S. & Knudsen, E. I. Experience-dependent plasticity in the inferior colliculus: a site for visual calibration of the neural representation of auditory space in the barn owl. *J. Neurosci.* **13**, 4589–4608 (1993).
- Hyde, P. S. & Knudsen, E. I. A topographic instructive signal guides the adjustment of the auditory space map in the optic tectum. *J. Neurosci.* (in the press).
- Feldman, D. E. & Knudsen, E. I. An anatomical basis for visual calibration of the auditory space map in the barn owl's midbrain. *J. Neurosci.* **17**, 6820–6837 (1997).
- DeBello, W. M., Feldman, D. E. & Knudsen, E. I. Adaptive axonal remodelling in the midbrain auditory space map. *J. Neurosci.* **21**, 3161–3174 (2001).
- Stein, B. E. & Meredith, M. A. *The Merging of the Senses* (MIT Press, Cambridge, Massachusetts, 1993).
- Knudsen, E. I. Auditory and visual maps of space in the optic tectum of the owl. *J. Neurosci.* **2**, 1177–1194 (1982).
- Middlebrooks, J. C. & Knudsen, E. I. A neural code for auditory space in the cat's superior colliculus. *J. Neurosci.* **4**, 2621–2634 (1984).
- King, A. J. & Hutchings, M. E. Spatial response properties of acoustically responsive neurons in the superior colliculus of the ferret: A map of auditory space. *J. Neurophysiol.* **57**, 596–624 (1987).
- King, A. J. & Carlile, S. Changes induced in the representation of auditory space in the superior colliculus by rearing ferrets with binocular eyelid suture. *Exp. Brain Res.* **94**, 444–455 (1993).
- Knudsen, E. I. & Brainard, M. S. Creating a unified representation of visual and auditory space in the brain. *Annu. Rev. Neurosci.* **18**, 19–44 (1995).
- Knudsen, E. I. & Knudsen, P. F. Space-mapped auditory projections from the inferior colliculus to the optic tectum in the barn owl (*Tyto alba*). *J. Comp. Neurol.* **218**, 187–196 (1983).
- Hyde, P. S. & Knudsen, E. I. Topographic projection from the optic tectum to the auditory space map in the inferior colliculus of the barn owl. *J. Comp. Neurol.* **421**, 146–160 (2000).
- Luksch, H., Gauger, B. & Wagner, H. A candidate pathway for a visual instructional signal to the barn owl's auditory system. *J. Neurosci.* **20**, RC70:1–4 (2000).
- Gold, J. I. & Knudsen, E. I. A site of auditory experience-dependent plasticity in the neural representation of auditory space in the barn owl's inferior colliculus. *J. Neurosci.* **20**, 3469–3486 (2000).
- Knudsen, E. I. & Mogdans, J. Vision-independent adjustment of unit tuning to sound localization cues in response to monaural occlusion in developing owl optic tectum. *J. Neurosci.* **12**, 3485–3493 (1992).
- Knudsen, E. I., Esterly, S. D. & du Lac, S. Stretched and upside-down maps of auditory space in the optic tectum of blind-reared owls: acoustic basis and behavioral correlates. *J. Neurosci.* **11**, 1727–1747 (1991).
- Brainard, M. S. & Knudsen, E. I. Sensitive periods for visual calibration of the auditory space map in the barn owl optic tectum. *J. Neurosci.* **18**, 3929–3942 (1998).
- Zheng, W. & Knudsen, E. I. Gabaergic inhibition antagonizes adaptive adjustment of the owl's auditory space map during the initial phase of plasticity. *J. Neurosci.* **21**, 4356–4365 (2001).
- Knudsen, E. I., Cohen, Y. E. & Masino, T. Characterization of a forebrain gaze field in the archistriatum of the barn owl: microstimulation and anatomical connections. *J. Neurosci.* **15**, 5139–5151 (1995).
- King, A. J., Hutchings, M. E., Moore, D. R. & Blakemore, C. Developmental plasticity in the visual and auditory representations in the mammalian superior colliculus. *Nature* **332**, 73–76 (1988).
- Withington-Wray, D. J., Binns, K. E., Dhanjal, S. S., Brickley, S. G. & Keating, M. J. The maturation of the superior collicular map of auditory space in the guinea pig is disrupted by developmental auditory deprivation. *Eur. J. Neurosci.* **2**, 693–703 (1990).
- Gold, J. I. & Knudsen, E. I. Hearing impairment induces frequency-specific adjustments in auditory spatial tuning in the optic tectum of young owls. *J. Neurophysiol.* **82**, 2197–2209 (1999).
- Gold, J. I. & Knudsen, E. I. Abnormal auditory experience induces frequency-specific adjustments in unit tuning for binaural localization cues in the optic tectum of juvenile owls. *J. Neurosci.* **20**, 862–877 (2000).
- Batista, A. P., Buneo, C. A., Snyder, L. H. & Andersen, R. A. Reach plans in eye-centered coordinates. *Science* **285**, 257–260 (1999).
- Cohen, Y. E. & Andersen, R. A. Reaches to sounds encoded in an eye-centered reference frame. *Neuron* **27**, 647–652 (2000).
- Troyer, T. W. & Doupe, A. J. An associational model of birdsong sensorimotor learning I. Efference copy and the learning of song syllables. *J. Neurophysiol.* **84**, 1204–1223 (2000).
- Klier, E. M., Wang, H. & Crawford, J. D. The superior colliculus encodes gaze commands in retinal coordinates. *Nature Neurosci.* **4**, 627–632 (2001).

**Supplementary Information** accompanies the paper on *Nature's* website (<http://www.nature.com>).

### Acknowledgements

We thank B. Newsome, J. Raymond, S. McConnell, J. Gold, B. Linkenhoker and Y. Gutfreund for comments on the manuscript. This work was supported by grants from the March of Dimes and the National Institute on Deafness and other Communication Disorders (NIDCD).

### Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.K. (e-mail: [eknudsen@stanford.edu](mailto:eknudsen@stanford.edu)).

# Stage-specific control of neuronal migration by somatostatin

Elina Yacubova & Hitoshi Komuro

Department of Neurosciences, Lerner Research Institute,  
The Cleveland Clinic Foundation, Cleveland, Ohio 44195, USA

Developing neurons transiently express somatostatin and its receptors<sup>1–3</sup>, but little is known about their function at these early stages. As we thought that endogenous somatostatin might control the migratory behaviour of immature neurons, we have examined the effects of somatostatin in cerebellar granule cells of early postnatal mice, because these cells express all five types of somatostatin receptors before the initiation of their migration<sup>4,5</sup>. Here we show that somatostatin has opposite and stage-specific effects on the migration of cerebellar granule cells. Activation of somatostatin receptors increases the rate of granule cell migration near their birthplace, but decreases the rate near their final destination. Furthermore, somatostatin enhances the size and frequency of spontaneous  $Ca^{2+}$  fluctuations in the early phase of migration, whereas it eliminates spike-like  $Ca^{2+}$  transients in the late phase. Somatostatin-induced changes at both early and late phases are reversed by a blockade of  $K^+$  channel activity. These results indicate that somatostatin may provide an essential cue for accelerating the movement of granule cells in the early phase and for terminating the movement in the late phase through altering intracellular  $Ca^{2+}$  concentrations and  $K^+$  channel activity.

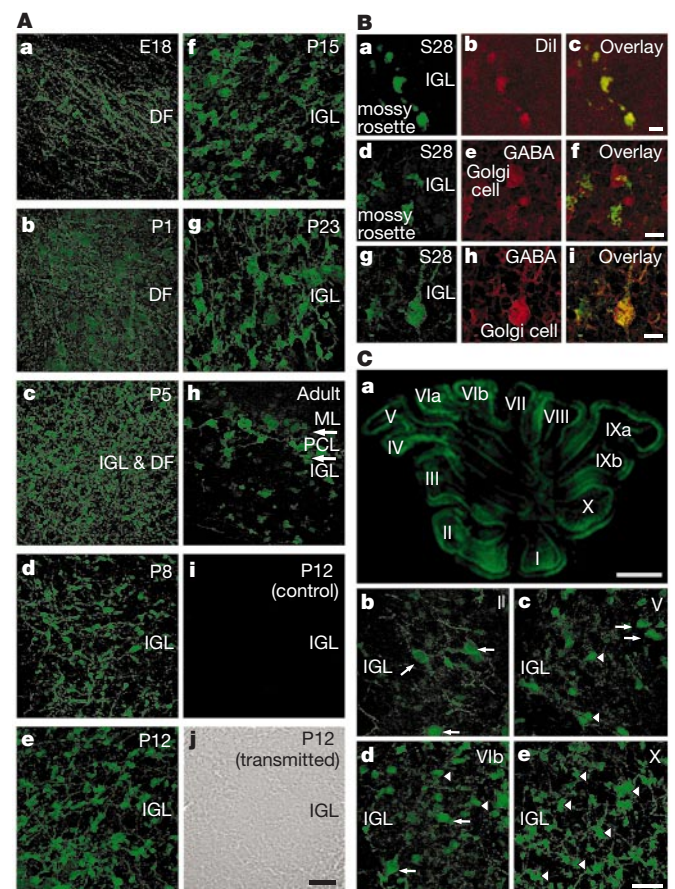
Somatostatin has two bioactive products, somatostatin-14 (SST-14) and somatostatin-28 (SST-28), which is a congener of SST-14 that is extended at the amino terminus<sup>6</sup>. Both peptides are produced in various proportions by different somatostatin cells and bind to all five somatostatin receptors<sup>6</sup>. During a period of granule cell migration, SST-14 is present in Purkinje cells, Golgi cells and climbing fibres<sup>2,7</sup>, but little is known about the expression of SST-28.

We therefore determined the onset and distribution of SST-28 in the developing cerebellum by using a polyclonal antibody against SST-28. We found that SST-28 was present in the differentiating field of the cerebellum as early as embryonic day 18 (E18; Fig. 1A). Thereafter, SST-28 concentrations increased gradually in the internal granular layer (IGL) and peaked during the first 2 weeks after birth. On the completion of neuronal migration, SST-28 concentrations decreased to those seen in adult. In the IGL, SST-28-positive cells were Golgi cells; these cells had larger cell bodies than granule cells and were GABA ( $\gamma$ -aminobutyric acid) positive (Fig. 1B). SST-28 positive fibres in the IGL were mossy fibre terminals; these fibres were GABA negative, and DiI (1,1'-diiododecyl-3,3,3',3'-tetramethylindocarbocyanine perchlorate)-labelled mossy rosettes were stained with the SST-28 antibody (Fig. 1B). Notably, each lobule showed a unique pattern of SST-28 expression (Fig. 1C). In lobule-I, for example, many Golgi cells contained SST-28, whereas mossy rosettes did not. In contrast, mossy rosettes were SST-28 positive in lobule-X, and SST-28-containing Golgi cells were rare. In lobules-V and -VIb, SST-28 was present evenly in both Golgi cells and mossy rosettes. These results, together with previous studies, show that during the period of granule cell movement, high concentrations of somatostatin (SST-14 or SST-28) are present along the migratory route and in the final destination of migrating granule cells.

We used acute cerebellar slice preparations to determine whether somatostatin controls the migratory behaviour of granule cells<sup>8–10</sup>. Time-lapse imaging of DiI-labelled cells showed that granule cells have three distinct modes of migration: (1) glia-independent tangential migration in the external granular layer (EGL); (2) glia-associated radial migration in the molecular layer; and (3) glia-independent radial migration in the IGL (Fig. 2a, b). The

addition of 1  $\mu$ M of SST-14 or SST-28 to the medium increased significantly the rate of cell movement in the EGL of lobule-V to 122% or 124% of the control rate, respectively (Fig. 2c). SST-14 or SST-28 slightly decreased the rate in the molecular layer to 90% or 88% of the control rate, respectively. Moreover, SST-14 or SST-28 significantly decreased the rate in the IGL to 57% or 53% of the control rate, respectively. These results show that somatostatin accelerates granule cell movement near the birthplace in the EGL, but significantly slows down the movement near their final destination in the IGL.

To determine the role of endogenous somatostatin in granule cell migration, we inhibited the activity of somatostatin receptors by applying a somatostatin antagonist, AC-178,335 (Ac-His-Phe-Ile-Arg-Trp-Phe-NH<sub>2</sub>, all D-amino acids), which effectively blocks the function of endogenous somatostatin *in vivo* and *in vitro*<sup>11</sup>. We examined the effects of AC-178,335 in granule cell migration in lobule-I, -V and -X, as these lobules showed a unique pattern of endogenous somatostatin (Fig. 1C). The addition of 10  $\mu$ M of AC-178,335 to the medium decreased significantly the rate of tangential migration of granule cells in the EGL to 76% (lobule-I), 74% (V) and 70% (X) of the control rate (Fig. 3). In contrast, AC-178,335 slightly increased the rate of radial migration in the molecular layer to 111% (I), 108% (V) and 109% (X) of the control rate. Notably,



**Figure 1** Expression of somatostatin (SST-28) in the cerebellum. **A**, **a–h**, SST-28 expression in lobule-X at different ages. **i**, **j**, Control images processed without primary antibody. **B**, **a–c**, Mossy rosettes stained with DiI are SST-28 positive at P5. **d–f**, SST-28 positive mossy terminals are GABA negative at P10. **g–i**, Golgi cells are SST-28 and GABA positive at P10. **C**, **a**, Sagittal section of P12 cerebellum stained with SST-28 antibody. **b–e**, Unique expression of SST-28 among four different lobules. Arrows and arrowheads represent Golgi cells and mossy rosettes, respectively. DF, differentiating field; IGL, internal granular layer; ML, molecular layer. Scale bars, 24  $\mu$ m (**A**, **j**); 2.5  $\mu$ m (**B**, **c**); 12  $\mu$ m (**B**, **f** and **i**); 700  $\mu$ m (**C**, **a**); 40  $\mu$ m (**C**, **e**).

AC-178,335 significantly increased the rate of radial migration in the IGL to 118% (I), 121% (V) and 129% (X) of the control rate. The application of AC-178,335 neither altered the morphology of granule cells nor changed the direction of cell movement. These results indicate that endogenous somatostatin differentially controls the migration of granule cells in the EGL, IGL and molecular layer, although the degree of alteration of the cell migration rate by endogenous somatostatin may vary slightly among each lobule.

To determine whether somatostatin acts directly on migrating granule cells or acts on other cells, which then indirectly influence granule cell migration, we examined the microexplant cultures of postnatal day 0 (P0) to P2 mouse cerebella<sup>12,13</sup>. In these cultures, most granule cells are derived from the EGL, as at the age of P0–P2 the IGL only contains small numbers of postmigratory granule cells<sup>14</sup>. Granule cells migrated without any contacts with other cells or processes for up to 2–3 days *in vitro* (Fig. 4A), similar to the time required for the cells *in vivo* to complete migration<sup>9,10</sup>. At 2 days *in vitro* (d.i.v.), granule cells started to express the  $\alpha 6$  subunit of GABA<sub>A</sub> receptors (Fig. 4B), suggesting that granule cells at 2 d.i.v. are in a similar stage of differentiation as the cells in the IGL<sup>15</sup>.

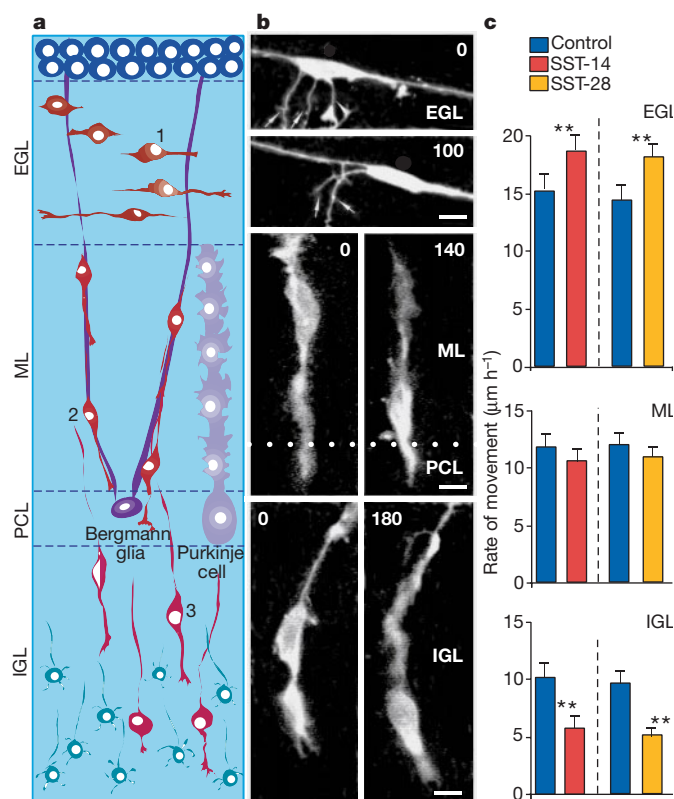
The application of 1  $\mu$ M of SST-14 or SST-28 failed to alter the cell shape, leading process length, direction or salutatory mode of granule cell migration at both 1 d.i.v. and 2 d.i.v. However, SST-14 or SST-28 significantly increased the rate of cell movement at 1 d.i.v. to 128% or 126% of the control rate, respectively, whereas SST-14 or SST-28 substantially decreased the rate at 2 d.i.v. to 53% or 58%, respectively (Fig. 4C). These results suggest that somatostatin

directly acts on migrating granule cells in a stage-specific manner. For the remaining experiments, we used SST-14 alone, because SST-14 showed slightly higher binding affinity to its receptors than did SST-28 (refs 16, 17).

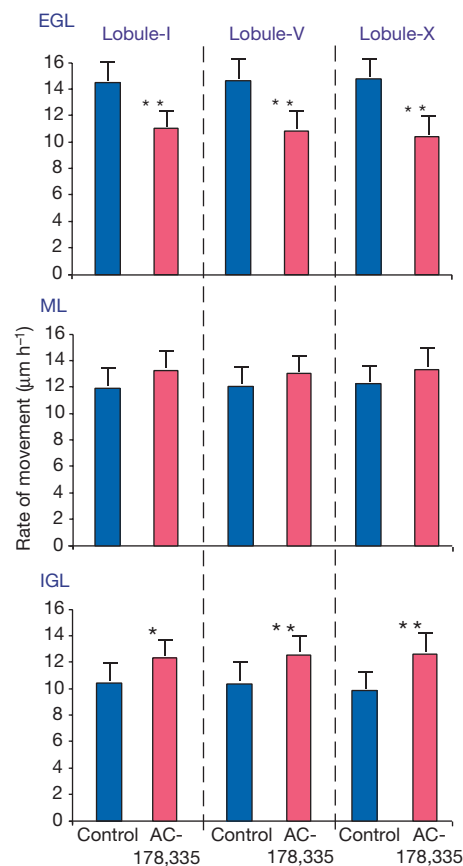
Our working hypothesis is that somatostatin controls granule cell movement by altering intracellular concentrations of  $\text{Ca}^{2+}$ . Somatostatin receptor activation leads to decrease  $\text{Ca}^{2+}$  currents through voltage-dependent  $\text{Ca}^{2+}$  channels in many cells<sup>18,19</sup>. The rate of granule cell migration depends on  $\text{Ca}^{2+}$  channel activity and  $\text{Ca}^{2+}$  current<sup>20,21</sup>, and is positively correlated with the size and frequency of  $\text{Ca}^{2+}$  fluctuations<sup>13</sup>. At 1 d.i.v., migrating granule cells showed small and irregular  $\text{Ca}^{2+}$  fluctuations (Fig. 4D, a). At 2 d.i.v.,  $\text{Ca}^{2+}$  fluctuations became larger and more regular (Fig. 4D, c). SST-14 (1  $\mu$ M) increased the size and frequency of  $\text{Ca}^{2+}$  fluctuations at 1 d.i.v., whereas SST-14 eliminated spike-like  $\text{Ca}^{2+}$  transients at 2 d.i.v. in 5 min after the application (Fig. 4D, a, c).

An increased rate of cell movement followed the enlargement of  $\text{Ca}^{2+}$  fluctuations at 1 d.i.v.; by contrast, the elimination of  $\text{Ca}^{2+}$  fluctuations at 2 d.i.v. decreased the rate (Fig. 4D, b, d). Furthermore, the  $\text{Ca}^{2+}$  chelator, BAPTA-AM, which clamps intracellular  $\text{Ca}^{2+}$  concentrations at low levels<sup>21</sup>, was applied to examine the involvement of  $\text{Ca}^{2+}$  fluctuations on somatostatin-induced changes. BAPTA-AM substantially decreased the rate of cell movement at both 1 d.i.v. and 2 d.i.v. regardless of the presence or absence of SST-14 (Fig. 4E). These results show that dynamic changes in intracellular  $\text{Ca}^{2+}$  concentrations are necessary for granule cell migration at both 1 and 2 d.i.v. The differential effects on intracellular  $\text{Ca}^{2+}$  concentrations might explain how somatostatin switches its effect on granule cell migration from acceleration at the early phase of the migration to slowdown at late phase of the migration.

We examined further how somatostatin regulates granule cell



**Figure 2** Role of somatostatin on granule cell migration in different cortical layers. **a**, Representation of granule cell migration from the external granular layer (EGL) to the internal granular layer (IGL): (1) tangential migration in the EGL; (2) Bergmann-glia-associated radial migration in the molecular layer (ML); (3) radial migration in the IGL. **b**, Migration of Dil-labelled granule cells in acute slices of P10 cerebellum. Arrows indicate vertical processes. Elapsed time (in minutes) is indicated in each photograph. Scale bars, 7  $\mu$ m. PCL, Purkinje cell layer. **c**, Effect of 1  $\mu$ M of SST-14 or SST-28 on granule cell migration in three cortical layers. Asterisks indicate statistical significance ( $P < 0.01$ ). Error bars indicate standard deviation.



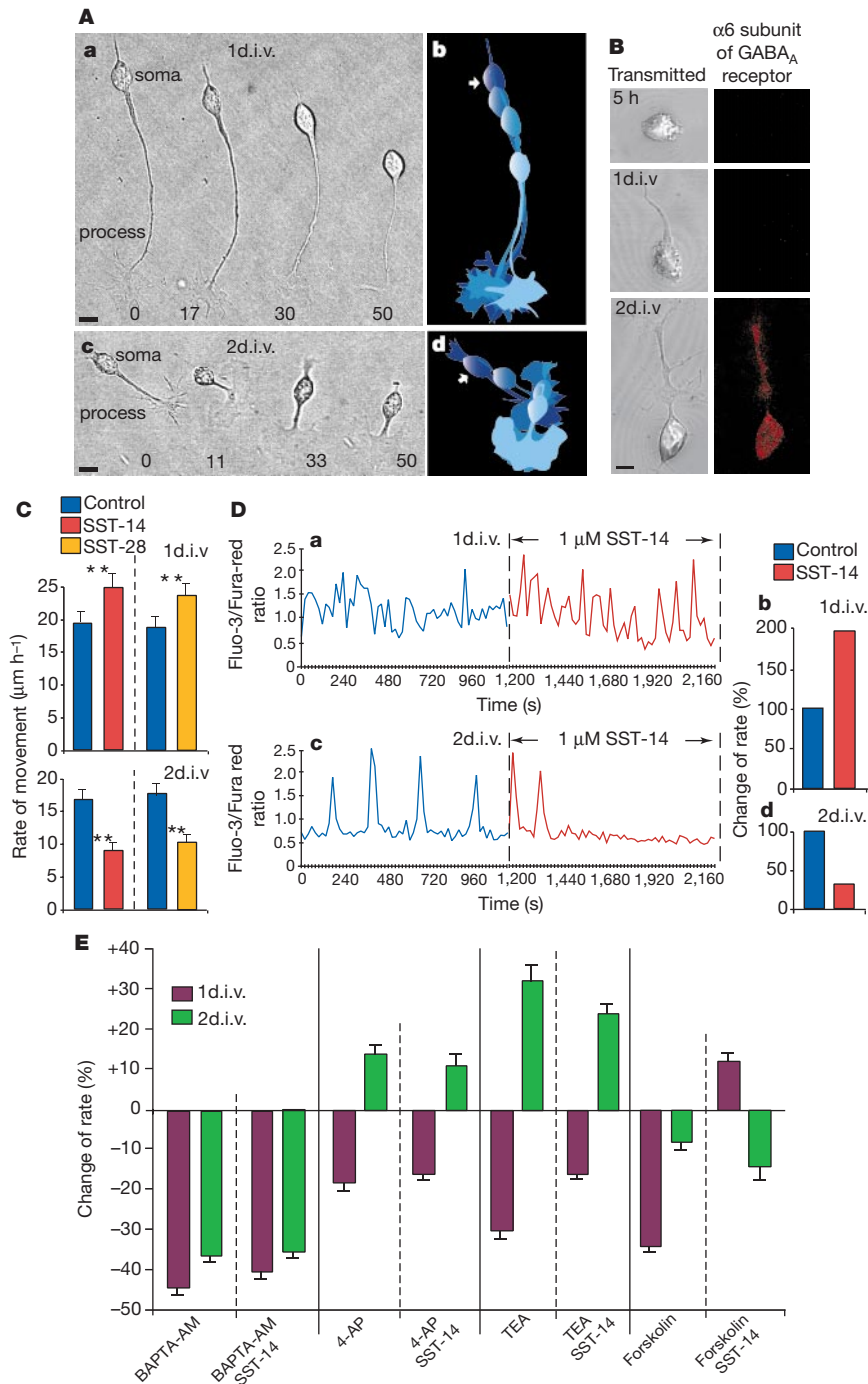
**Figure 3** Effect of a somatostatin antagonist, AC-178,335 (10  $\mu$ M), on granule cell migration. Single ( $P < 0.05$ ) and double ( $P < 0.01$ ) asterisks indicate statistical significance. Error bars indicate standard deviation.



movement in a stage-specific manner. First, the application of thapsigargin ( $5 \mu\text{M}$ ), which is a  $\text{Ca}^{2+}$ -ATPase inhibitor that depletes intracellular  $\text{Ca}^{2+}$  stores<sup>13</sup>, substantially decreased the rate of cell movement to  $41 \pm 4\%$  (mean  $\pm$  s.d.,  $n = 32$ ), and the size of  $\text{Ca}^{2+}$  fluctuations to  $56 \pm 5\%$  ( $n = 32$ ) of the control values at 1 d.i.v.. However, the effect of thapsigargin was less at 2 d.i.v.: the rate of cell movement and the size of  $\text{Ca}^{2+}$  fluctuations were  $87 \pm 7\%$  ( $n = 31$ ) and  $90 \pm 6\%$  ( $n = 31$ ) of the control values, respectively. Second, the reduction of  $\text{Ca}^{2+}$  influxes by lowering extracellular  $\text{Ca}^{2+}$

concentrations from 1.8 mM to 0.1 mM decreased both the rate of cell movement and the size of  $\text{Ca}^{2+}$  fluctuations at 1 d.i.v. to  $68 \pm 4\%$  ( $n = 32$ ) and  $74 \pm 5\%$  ( $n = 32$ ) of the control values, respectively. The effects of low extracellular  $\text{Ca}^{2+}$  concentrations were more prominent at 2 d.i.v.: the rate of cell movement and the size of  $\text{Ca}^{2+}$  fluctuations were  $23 \pm 3\%$  ( $n = 36$ ) and  $35 \pm 5\%$  ( $n = 36$ ) of the control values, respectively.

Third, we determined the role of  $\text{K}^+$  channels on somatostatin-induced changes in cell movement, because somatostatin enhances



**Figure 4** Somatostatin alters granule cell migration in the microexplant cultures. **A, a, c**, Time-lapse images of granule cells. Elapsed time (in minutes) is indicated on bottom of each photograph. **b, d**, Superimposed images of **a** and **c**. Arrows indicate first images. Scale bars, 8  $\mu\text{m}$ . **B**, Expression of  $\alpha 6$  subunit of  $\text{GABA}_A$  receptors in granule cells. Scale bar, 3  $\mu\text{m}$ . **C**, Effect of 1  $\mu\text{M}$  of SST-14 or SST-28 on rate of granule cell migration.

Asterisks indicate statistical significance ( $P < 0.01$ ). **D, a–d**, Changes in intracellular  $\text{Ca}^{2+}$  fluctuations and speed of granule cell movement by SST-14 (1  $\mu\text{M}$ ). **E**, Effect of BAPTA-AM (10  $\mu\text{M}$ ), 4-AP (2 mM), TEA (2 mM), or forskolin (30  $\mu\text{M}$ ) on SST-14-induced changes in granule cell migration. Error bars in **C** and **E** indicate standard deviation.

activity of several types of  $K^+$  channels in many cells<sup>16,17,22</sup>. The blockade of  $K^+$  channel activity by tetraethylammonium chloride (TEA) or 4-aminopyridine (4-AP) completely reversed the somatostatin-induced changes in cell movement at both 1 d.i.v. and 2 d.i.v. (Fig. 4E). Fourth, the elevations of intracellular cAMP concentrations by forskolin failed to alter somatostatin-induced changes in cell movement, although forskolin decreased the rate of cell movement at both 1 d.i.v. and 2 d.i.v. in the absence of SST-14 (Fig. 4E). These results indicate that granule cell movement mainly depends on  $Ca^{2+}$  release from intracellular  $Ca^{2+}$  stores at 1 d.i.v., and on  $Ca^{2+}$  influxes across the plasma membrane at 2 d.i.v.. Somatostatin might control granule cell migration by increasing  $K^+$  channel activity. Potassium channels could act together with  $Ca^{2+}$  channels to control migration of immature neurons<sup>23</sup>. The alteration of  $K^+$  channel activity might differentially affect  $Ca^{2+}$  channel activity of granule cells at 1 d.i.v. and 2 d.i.v., because their resting membrane potential, which negatively shifts during a culture period<sup>24</sup>, is near or within activation ranges of  $Ca^{2+}$  channels<sup>25</sup>. Furthermore, granule cells at 1 d.i.v. might use different extrinsic and intrinsic signals to control their movement than those used at 2 d.i.v., which would account for the differential effects of somatostatin on their migration.

The issue of whether somatostatin and its receptors have a role in the development of brain has not been resolved for more than a decade. In this series of experiments, we have shown a transient but specific role of somatostatin and its receptors in controlling the rate of granule cell migration. Changes in intracellular  $Ca^{2+}$  concentrations and  $K^+$  channel activity are probably involved in somatostatin-induced alteration of granule cell migration, but the cellular and molecular mechanisms of its stage-specific control of cell movement remain to be examined. Somatostatin might control migration of immature neurons other than cerebellar granule cells, specifically in cerebrum<sup>26</sup> and spinal cord<sup>3</sup>. In these brain areas, neurons expressing high concentrations of somatostatin receptors migrate along somatostatin-containing neurons. Moreover, early expression of the somatostatin and its receptors may be an essential prerequisite to the initiation of migration, as somatostatin enhances neurite outgrowth<sup>27</sup> and potentiates expression of  $K^+$  and  $Ca^{2+}$  channels<sup>28</sup>. □

## Methods

### Immunohistochemistry and DiI labelling

Both genders of CD-1 mice, from 18-day-old embryo to adult, were used in this study. After an anaesthetic overdose, animals were perfused transcardially with PBS pH 7.4, followed by 4% paraformaldehyde in PBS (PFA). Brains were fixed in PFA overnight, transferred to 30% sucrose solution in PBS for an additional 24 h, and 30- $\mu$ m sections were prepared on a cryostat. We blocked sections in 10% serum, 0.1% Triton-X, and incubated them overnight in anti-somatostatin 28 antibody 1:800 and/or anti-GABA antibody 1:800 (Cemicon International). Anti-somatostatin 28 antibody only recognizes SST-28. Sections were washed and incubated in fluorescein-conjugated or Texas-red-conjugated secondary antibody (Jackson Immuno Research). Cultured granule cells were fixed in PFA for 20 min and incubated with anti-GABA $\alpha$ 6 subunit antibody 1:1,000 (Cemicon International) overnight. To label mossy terminals, we inserted small DiI crystals into pons nuclei of 5-day-old mice.

### Granule cell migration in acute slice preparations

Methods for monitoring granule cell migration in cerebellar slice preparations have been described<sup>8–10</sup>. Briefly, cerebella obtained from P10 (CD-1) mice were sectioned transversely or sagittally into 300- $\mu$ m slices. To label granule cells, cerebellar slices were incubated for 3 min in a fluorescent lipophilic carbocyanine dye (DiI; 7  $\mu$ g ml<sup>-1</sup>) (Molecular Probes), which was added to the culture medium. The culture medium comprised minimum essential medium (Gibco) supplemented with 30 mM glucose, 1.8 mM glutamine, 24 mM NaHCO<sub>3</sub>. We used a confocal microscope (Leica) to visualize DiI-labelled granule cells in the slices. The tissue was illuminated with a 488-nm wavelength light from an argon laser, and fluorescence emission was detected at 530  $\pm$  15 nm. Images of migrating granule cells were collected every minute. The rate of movement was calculated from each 60-min recording before and after addition of somatostatin or somatostatin antagonist (AC-178,335). Each column in Figs 2C and 3 represents the average rate of cell movement obtained from more than 30 migrating cells.

### Microexplant culture

Methods for microexplant cultures of early postnatal mouse cerebella have been

described<sup>12,13</sup>. Small pieces of cerebellum obtained from P0–P2 CD-1 mice were placed on poly-L-lysine/laminin-coated coverslips. Each coverslip was transferred into a 35-mm Petri dish and put in a CO<sub>2</sub> incubator (37°C, 95% air, 5% CO<sub>2</sub>). The incubation medium comprised minimum essential medium supplemented with 5% fetal calf serum, 30 mM glucose, 1.8 mM glutamine, 24 mM NaHCO<sub>3</sub>, 90 U ml<sup>-1</sup> penicillin and 90  $\mu$ g ml<sup>-1</sup> streptomycin. One or two days after transplantation, coverslips were transferred into the chamber of a micro-incubator attached to the stage of an inverted microscope. We used a confocal microscope to monitor the changes in cell shape and migratory behaviour of granule cells before and after application of somatostatin. Transmitted images at 488 nm were collected every 20 s for up to 3 h. Thapsigargin (5  $\mu$ M), 4-AP (2 mM), TEA (2 mM), forskolin (30  $\mu$ M) or BAPTA-AM (10  $\mu$ M) was added to the medium in the presence or absence of SST-14 (1  $\mu$ M) after a 60-min control observation. We evaluated the effects of each agent by dividing distance travelled in the first 60 min in the absence of each reagent by distance travelled in the 60 min after its application. Each column in Fig. 4C and E represents the average rate of cell movement obtained from more than 30 migrating cells.

### Ca<sup>2+</sup> measurements

Methods for Ca<sup>2+</sup> imaging from migrating neurons in the microexplant culture have been described<sup>13</sup>. Briefly, 1–2 d after transplantation, granule cells were co-loaded with the cell-permeant, acetoxymethyl ester form of 3  $\mu$ M Fluo-3 and 4  $\mu$ M Fura-red (Molecular Probes) diluted in culture medium for 30 min at 37°C. To determine whether the loading of the two Ca<sup>2+</sup> indicator dyes affects cell migration, we examined changes in granule cell migration before and after loading 3  $\mu$ M Fluo-3 and 4  $\mu$ M Fura-red. The loading of these Ca<sup>2+</sup> indicator dyes did not cause any significant noxious effects on granule cell migration. For example, the rate of cell movement was 19.3  $\pm$  1.5  $\mu$ m h<sup>-1</sup> ( $n$  = 51) at 1 d.i.v. and 17.1  $\pm$  2.0  $\mu$ m h<sup>-1</sup> ( $n$  = 44) at 2 d.i.v. before loading, and 18.8  $\pm$  1.4  $\mu$ m h<sup>-1</sup> ( $n$  = 51) at 1 d.i.v. and 16.8  $\pm$  1.4  $\mu$ m h<sup>-1</sup> ( $n$  = 44) at 2 d.i.v. after loading. We used a confocal microscope to examine the intracellular Ca<sup>2+</sup> concentrations and the rate of cell movement. The cells were illuminated with 488-nm wavelength light from an argon laser. Fluorescence images for ratiometric Ca<sup>2+</sup> measurements (at 530  $\pm$  15 nm for Fluo-3; at 640  $\pm$  15 nm for Fura-red) and transmitted images for monitoring cell movement were collected simultaneously every 20 s for up to 3 h. Fluo-3/Fura-red ratio images were calculated by dividing pixel values of Fluo-3 fluorescence intensity by pixel values of Fura-red fluorescence intensity.

Received 29 May; accepted 30 October 2001.

- Gonzalez, B. J. *et al.* Transient expression of somatostatin receptors in the rat cerebellum during development. *Dev. Brain Res.* **40**, 154–157 (1988).
- Inagaki, S. *et al.* In situ hybridization analysis of the somatostatin-containing neuron system in developing cerebellum of rats. *Mol. Brain Res.* **6**, 289–295 (1989).
- Maubert, E. *et al.* Developmental patterns of somatostatin-receptors and somatostatin-immuno-reactivity during early neurogenesis in the rat. *Neuroscience* **62**, 317–325 (1994).
- Thoss, V. S., Duc, D. & Hoyer, D. Somatostatin receptors in the developing rat brain. *Eur. J. Pharmacol.* **297**, 145–155 (1996).
- Violet, C. *et al.* Differential expression of multiple somatostatin receptors in the rat cerebellum during development. *J. Neurochem.* **68**, 2263–2272 (1997).
- Schinder, M., Humphrey, P. P. A. & Emson, P. C. Somatostatin receptors in the central nervous system. *Prog. Neurobiol.* **50**, 9–47 (1996).
- Villar, M. J., Hofkelt, T. & Brown, J. C. Somatostatin expression in the cerebellar cortex during postnatal development. *Anat. Embryol.* **179**, 257–267 (1989).
- Komuro, H. & Rakic, P. Dynamics of granule cell migration: A confocal microscopic study in acute cerebellar slice preparations. *J. Neurosci.* **15**, 1110–1120 (1995).
- Komuro, H. & Rakic, P. Distinct modes of neuronal migration in different domains of developing cerebellar cortex. *J. Neurosci.* **18**, 1478–1490 (1998).
- Komuro, H., Yacubova, E., Yacubova, E. & Rakic, P. Mode and tempo of tangential cell migration in the cerebellar external granular layer. *J. Neurosci.* **21**, 527–540 (2001).
- Baumbach, W. R. *et al.* A linear hexapeptide somatostatin antagonist blocks somatostatin activity *in vitro* and influences growth hormone release in rats. *Mol. Pharmacol.* **54**, 864–873 (1998).
- Nagata, I. & Nakatsuji, N. Granule cell behavior on laminin in cerebellar microexplant cultures. *Dev. Brain Res.* **52**, 63–73 (1990).
- Komuro, H. & Rakic, P. Intracellular Ca<sup>2+</sup> fluctuations modulate the rate of neuronal migration. *Neuron* **17**, 275–285 (1996).
- Miale, I. L. & Sidman, R. L. An autoradiographic analysis of histogenesis in the mouse cerebellum. *Exp. Neurol.* **4**, 277–296 (1961).
- Kuhar, S. G. *et al.* Changing patterns of gene expression define four stages of cerebellar granule neuron differentiation. *Development* **117**, 97–104 (1993).
- Reisine, T. & Bell, G. I. Molecular biology of somatostatin receptors. *Endocr. Rev.* **16**, 427–442 (1995).
- Patel, Y. C. Molecular pharmacology of somatostatin receptor subtypes. *J. Endocrinol. Invest.* **20**, 348–367 (1997).
- Ikedo, S. R. & Schofield, G. G. Somatostatin blocks a calcium current in rat sympathetic ganglion neuron. *J. Physiol. (Lond.)* **409**, 221–240 (1989).
- Gonzalez, B. J. *et al.* Somatostatin receptors are expressed by immature cerebellar granule cells: Evidence for a direct inhibitory effect of somatostatin on neuroblast activity. *Proc. Natl Acad. Sci. USA* **89**, 9627–9631 (1992).
- Komuro, H. & Rakic, P. Selective role of N-type calcium channels in neuronal migration. *Science* **257**, 806–809 (1992).
- Komuro, H. & Rakic, P. Modulation of neuronal migration by NMDA receptors. *Science* **260**, 95–97 (1993).
- Sims, S. M., Lussier, B. T. & Kraicer, J. Somatostatin activates an inwardly rectifying K<sup>+</sup> conductance in freshly dispersed rat somatotropes. *J. Physiol. (Lond.)* **441**, 615–637 (1991).
- Hendriks, R., Morest, D. K. & Kaczmarek, L. K. Role of neuronal cell migration for high-threshold potassium currents in the chicken hindbrain. *J. Neurosci. Res.* **58**, 805–814 (1999).
- Shibata, R. *et al.* A-type K<sup>+</sup> current mediated by the Kv4 channel regulates the generation of action potential in developing cerebellar granule cells. *J. Neurosci.* **20**, 4145–4155 (2000).

25. D'Angelo, E., De Filippi, G., Rossi, P. & Taglietti, V. Synaptic activation of  $\text{Ca}^{2+}$  action potentials in immature rat cerebellar granule cells *in situ*. *J. Neurophysiol.* **78**, 1631–1642 (1997).
26. Bologna, E. & Leroux, P. Identification of multiple somatostatin receptors in the rat somatosensory cortex during development. *J. Comp. Neurol.* **420**, 466–480 (2000).
27. Taniwaki, T. & Schwartz, J. P. Somatostatin enhances neurofilament expression and neurite outgrowth in cultured rat cerebellar granule cells. *Dev. Brain Res.* **88**, 109–116 (1995).
28. Rhie, D.-J. *et al.* Somatostatin potentiates voltage-dependent  $\text{K}^+$  and  $\text{Ca}^{2+}$  channels expression induced by nerve growth factor in PC12 cells. *Dev. Brain Res.* **112**, 267–274 (1999).

#### Acknowledgements

We thank K. Wikler, M. Perin and P. Rakic for critically reading the manuscript. We also thank E. Yacubova and J. Neffenger for technical help and manuscript preparation. This work was supported by the Cleveland Clinic Foundation (H.K.).

#### Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.K. (e-mail: komuroh@ccf.org).

## Inter-receptor communication through arrays of bacterial chemoreceptors

Jason E. Gestwicki & Laura L. Kiessling

Departments of Chemistry and Biochemistry, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA

The sensing mechanisms of chemotactic bacteria allow them to respond sensitively to stimuli. *Escherichia coli*, for example, respond to changes in chemoattractant concentration of less than 10% over a range spanning six orders of magnitude<sup>1,2</sup>. Sensitivity over this range depends on a nonlinear relationship between ligand concentration and output response<sup>3</sup>. At low ligand concentrations, substantial amplification of the chemotactic signal is required; however, the mechanism responsible for this amplification remains unclear. Here we demonstrate that inter-receptor communication within a lattice<sup>4,5</sup> acts to amplify and integrate sensory information. Synthetic multivalent ligands that interact through the low-abundance, galactose-sensing receptor Trg stabilize large clusters of chemoreceptors and markedly enhance signal output from these enforced clusters. On treatment with multivalent ligands, the response to the attractant serine is amplified by at least 100-fold. This amplification requires a full complement of chemoreceptors; deletion of the aspartate (Tar) or dipeptide (Tap) receptors diminishes the amplification of the serine response. These results demonstrate that the entire array is involved in sensing. This mode of information exchange has general implications for the processing of signals by cellular receptors.

A family of at least four chemoreceptors (methyl-accepting chemotaxis proteins, MCPs) mediates metabolism-independent chemotactic responses in *E. coli*<sup>6,7</sup>. The high-abundance MCPs (Tsr and Tar) mediate responses to serine and aspartate, respectively; the low-abundance MCPs (Tap and Trg) are present at only 10% of the concentration of Tsr<sup>8,9</sup>. Responses to dipeptides are transmitted through Tap. Trg mediates responses to galactose and ribose by means of their respective periplasmic binding proteins. The MCPs transmit information through a two-component signalling cascade. These transmembrane proteins exist as dimers that complex to the cytoplasmic kinase CheA and the adapter protein CheW<sup>10,11</sup>. Ligand binding to the MCPs, either alone or together

with one of the periplasmic binding proteins, promotes a conformational change in the receptor that influences the activity of CheA<sup>12,13</sup>. CheA transfers a phosphoryl group to CheY, the response regulator that ultimately controls flagellar rotation and locomotion. When the attractant concentration remains stable, *E. coli* adapt through a process that involves methylation of glutamate residues in the cytoplasmic domains of the MCPs. The MCPs and other components of the signalling cascade localize to the cell poles of *E. coli*<sup>14</sup>. This pattern of MCP localization is conserved in other bacteria and in *Archaea*<sup>15</sup>.

Recent models implicate inter-receptor communication within chemoreceptor arrays in signal amplification<sup>4,5</sup>. Such a mechanism would complement findings indicating that inter-receptor transfer of methyl groups occurs during adaptation to chemotactic stimuli<sup>16</sup>. These results and others<sup>17</sup> support inter- and intradimer cross-talk during adaptation. With regard to signal amplification, however, the evidence supporting a role for chemoreceptor arrays is indirect. One proposal is that inter-receptor communication occurs between homodimers of one receptor-type<sup>18,19</sup>. This homodimer–homodimer communication may be mediated by transmission of ligand-induced conformational changes facilitated by receptor–receptor contact<sup>13,20</sup>. However, it is unclear what mechanisms of inter-receptor communication function during excitatory chemotactic responses in the entire MCP array. In lattice models of chemoreceptor function, the proximity of MCPs is a principal determinant of signalling<sup>4,5</sup>. We proposed, therefore, that the stabilization of receptor–receptor contacts would reveal the importance of this parameter for the amplification of chemotactic signals.

To create reagents that stabilize inter-receptor interactions, we synthesized multivalent derivatives of the chemoattractant galactose by ring-opening metathesis polymerization (ROMP) (Fig. 1a). ROMP allows the generation of multivalent ligands with distinct valencies<sup>21</sup>. Ligands of sufficient length (approximately 25 monomer units) would be expected to possess the ability to bind multiple MCPs simultaneously<sup>22</sup>. Galactose and synthetic galactose-bearing ligands<sup>23</sup> activate Trg through interaction with the periplasmic binding protein, glucose/galactose-binding protein (GGBP). We anticipated that the multivalent chemoattractants would stabilize inter-receptor contacts by forming complexes containing multiple copies of Trg, and thus could be used to elucidate the role of receptor position in signal amplification.

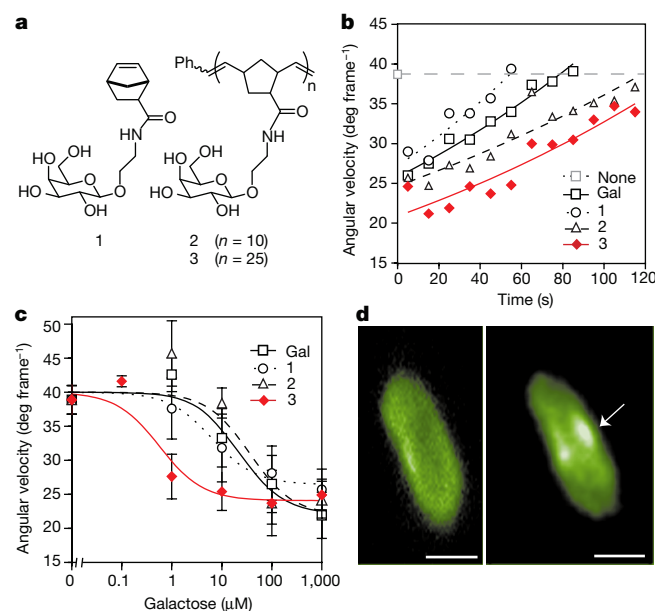
To investigate the impact of ligand valency on bacterial behaviour, we conducted motion analysis. The mean angular velocity of a bacterial population decreases in response to chemoattractants<sup>24</sup>. As expected<sup>22</sup>, angular velocity decreased when galactose or galactose derivatives 1–3 were introduced to cells of the chemotactically active *E. coli* strain at 100  $\mu\text{M}$  galactose (all concentrations were based on the total amount of galactose) (Fig. 1b). The magnitude of the response depended on the concentration and valency of the attractant (Fig. 1c). The adaptation time—the period before the population returns to pre-stimulus angular velocities—was also influenced by these factors (see Supplementary Information). Notably, multivalent ligand 3 was the most potent chemoattractant. This attractant was 100- to 1,000-fold more potent than its monovalent counterpart ligand 1. In contrast, multivalent ligand 2 did not have a significantly different chemotactic activity compared to monovalent ligand 1 or galactose at most concentrations. Molecular modelling studies indicate that multivalent ligand 2 is of insufficient length to cluster Trg<sup>23</sup>. These data suggest that the differences in chemotactic activity between ligands 3 and 1 are due to the ability of ligand 3 to stabilize MCP clusters.

To test whether the synthetic chemoattractants alter MCP clustering, we used fluorescence microscopy to monitor receptor position<sup>22</sup>. MCPs distribute randomly in *E. coli* RP1078 (ref. 25), a strain that lacks functional CheW<sup>14,22</sup>. Multivalent ligand 3, but not monovalent ligand 1, induces chemoreceptor clustering in this *cheW* mutant (Fig. 1d). The patterns of MCP distribution in cells



treated with multivalent ligand 2 were indistinguishable from those treated with monovalent ligand 1 (data not shown). Thus, only ligand 3, which is predicted to interact through multiple copies of Trg, promotes clustering of the chemoreceptors<sup>22</sup>.

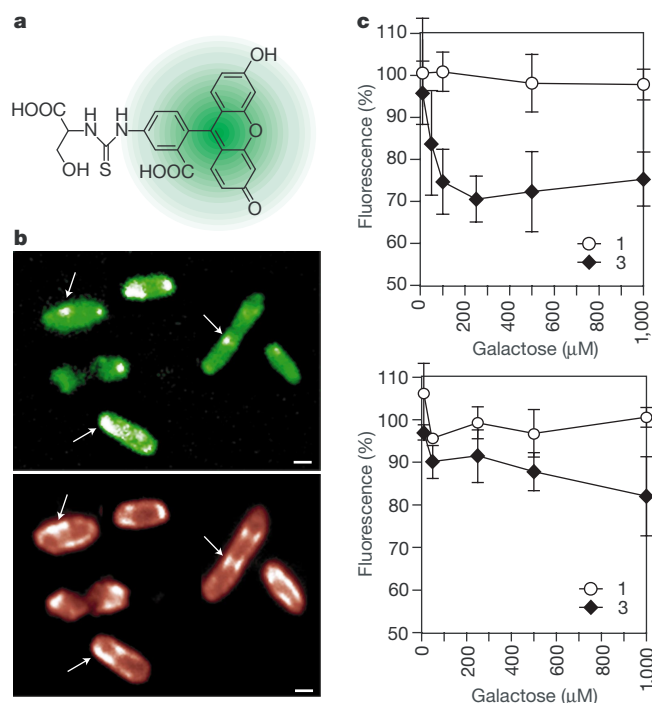
The chemoreceptor clusters that formed in the presence of multivalent ligand 3 were larger than would be expected for those only composed of Trg (approximately 150 copies per cell)<sup>8</sup>. We propose that these clusters are composed of heterogeneous collections of MCPs. The cytoplasmic and transmembrane domains that mediate inter-MCP contact are homologous (>50% sequence identity). Heterologous MCP clustering may therefore be mediated through Trg–MCP interactions, which are encouraged by multivalent ligand 3. Because Tsr is the MCP with the highest cellular concentration (roughly 3,000 copies per cell)<sup>9</sup>, it should constitute a large percentage of the receptors in clusters stabilized by multivalent ligand 3. Accordingly, we created a fluorescent ligand for the chemoreceptor Tsr (serine-fluorescein (Ser-fl); Fig. 2a) to probe Tsr location. When the *E. coli cheW* mutant was pre-treated with galactose derivatives 1 or 2 before Ser-fl addition, no change in MCP clustering was observed (see Supplementary Information). In contrast, when multivalent ligand 3 was added, the fluorescence due to Ser-fl was localized to clusters (Fig. 2b, top panel). These clusters colocalized with the MCPs (Fig. 2b, bottom panel). These data suggest that ligands for Trg (multivalent ligand 3) and Tsr (Ser-fl) bind to clusters of MCPs, indicating that both Trg and Tsr are included in the clusters reinforced by multivalent ligand 3.



**Figure 1** Effects of chemoattractants on the behaviour of *E. coli* and the localization of MCPs. **a**, Chemical structures of synthetic galactose-bearing monomer 1 and multivalent chemoattractants 2 and 3 (ref. 22). Ligand valency ( $n$ ) represents the degree of polymerization, which was determined by integration of the <sup>1</sup>H nuclear magnetic resonance (NMR) spectrum. **b**, Plot of average angular velocity versus time for *E. coli* AW405 treated with chemoattractant (galactose or ligands 1–3). Motion analysis was performed as described previously (see Supplementary Information)<sup>22,24</sup>. The ligand concentration (100  $\mu$ M) is given as the total concentration of galactose or galactose residues. Results are the average of at least four experiments performed in triplicate. **c**, Angular velocity of *E. coli* treated with various concentrations of chemoattractant. Data points represent the average angular velocity for the first 15 s after addition of attractant or buffer. **d**, Fluorescence micrographs of *cheWE. coli* RP1078 cells treated with galactose-bearing ligand 1 (left) or 3 (right). Microscopy was performed using fluorescein-labelled anti-MCP antibodies, as described<sup>22</sup>. A sample cluster is highlighted with an arrow. The average size of the clusters observed ( $10,000 \pm 3,500$  nm<sup>2</sup>) suggests that they contain MCPs other than Trg. The theoretical maximum size of a cluster containing 150 copies of Trg<sup>8,9</sup> is approximately 2,900 nm<sup>2</sup> (ref. 20). Scale bar, 0.5  $\mu$ m.

Fluorescence and electron microscopy experiments indicate the MCPs are localized<sup>14,15</sup>; however, these methods cannot discern whether the distances between receptors are appropriate for inter-receptor communication. We tested whether Ser-fl could serve as a sensitive probe of changes in Tsr clustering by comparing the fluorescence emission intensity of cells treated with Ser-fl in the presence of monovalent ligand 1 or multivalent ligand 3. The fluorescence intensity due to Ser-fl binding to either *cheW* mutants (Fig. 2c, top panel) or wild-type cells (Fig. 2c, bottom panel) was unaffected by pre-treatment with monovalent ligand 1. In contrast, on pre-treatment with multivalent ligand 3, the fluorescence intensity decreased 25–30% in the *cheW* mutant (Fig. 2c, top panel) and 15–20% in wild-type cells (Fig. 2c, bottom panel). These results are consistent with fluorescence self-quenching resulting from the clustering of Ser-fl binding sites. Our observations suggest that multivalent ligand 3 serves to stabilize inter-receptor contacts in wild-type cells, albeit to a lesser extent than in *cheW* mutants.

If inter-receptor communication is critical for signal amplification, responses mediated by Tsr should be enhanced in the presence of a ligand that increases MCP proximity. To test this hypothesis, wild-type *E. coli* were treated with ligands 1–3 and then allowed to adapt. After full adaptation, serine was added, and the angular velocity was measured. Pre-treatment with multivalent ligand 3, but not monovalent ligand 1 or multivalent ligand 2, amplified the response of the bacteria to serine (Fig. 3a). The serine concentration required to achieve a behavioural response was decreased at least 100-fold by pre-treatment with multivalent ligand 3. Similarly, responses to aspartate were augmented by pre-treatment with



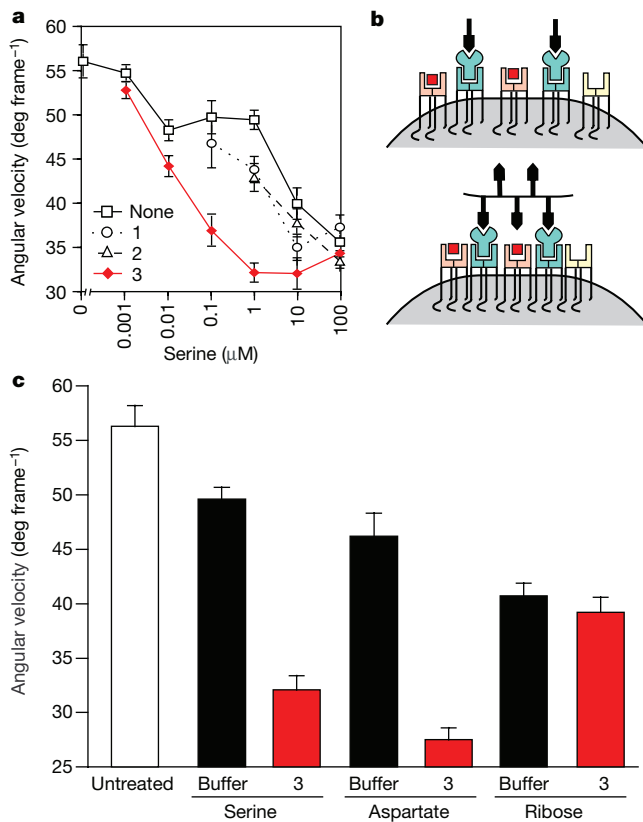
**Figure 2** Tsr is included in clusters formed on treatment with galactose-bearing ligand 3. **a**, Chemical structure of the serine-fluorescein conjugate (Ser-fl). Ser-fl served as a chemoattractant in wild-type AW405 but not Tsr mutant AW518 cells (see Supplementary Information). **b**, Treatment with 500  $\mu$ M galactose-bearing ligand 3 causes patching of Ser-fl binding sites (top). Patches (arrows) occur in both polar and non-polar locations and correspond to sites of anti-MCP antibody binding (bottom). Microscopy was performed as described previously<sup>22</sup>. Scale bar, 0.5  $\mu$ m. **c**, Fluorescence intensity of *cheW* mutant RP1078 (top panel) or wild-type AW405 (bottom panel) cells after treatment with ligands 1 or 3, followed by 30  $\mu$ M Ser-fl. The concentration of ligand 1 or 3 is reported as the concentration of galactose residues. Values ( $\pm$  s.e.) are the average of three independent experiments.

multivalent ligand 3 (Fig. 3c). To test whether this enhancement arises from attractant binding to clustered receptors, ribose—which also signals through Trg<sup>2</sup>—was added as a second attractant. If ribose displaces multivalent ligand 3, no signal enhancement should be observed. Even if ligand 3 remains bound, the low copy number of Trg should ensure that the clusters contain little if any unoccupied Trg. Accordingly, when ribose was the second attractant (Fig. 3c), no potentiation of response was observed. These results strongly support a role for inter-receptor communication in chemotactic signal amplification.

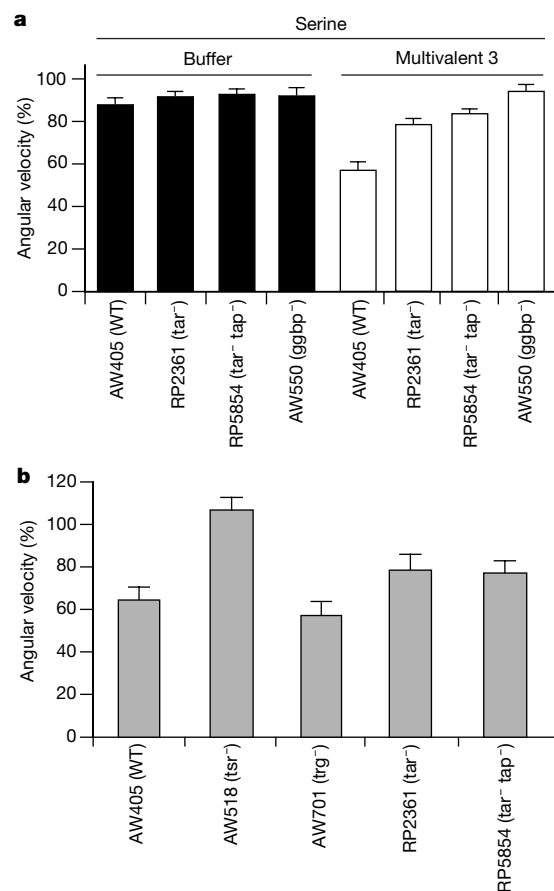
Next, we explored the contribution of various chemotactic signalling components to inter-receptor communication by analysing the responses of strains lacking one or more principal constituent. Multivalent ligand 3 had no effect on the responses to serine in a *ggbp* (*E. coli* AW550) or *trg* (*E. coli* AW701) mutant (see Supplementary Information). A *tsr* mutant (*E. coli* AW518)<sup>26</sup> also failed to respond to the combination of ligand 3 and serine, indicating that the serine receptor was required for amplification

of the chemotactic response (see Supplementary Information). Deletion of Tar (*E. coli* RP2361) decreased the potentiation of serine responses from 36% to 15% (Fig. 4a). When both Tar and Tap (*E. coli* RP5854) are deleted, the potentiation decreases by an additional 5%. Thus, a full complement of MCPs is required for maximal potentiation of serine responses—even receptors that do not bind chemoattractant contribute to the response.

We proposed that multivalent ligand 3, by stabilizing inter-receptor contact, would amplify any tendency of MCPs to perform as a lattice during sensory stimulation. Given the enhancement of the serine response observed in the presence of ligand 3 and its dependence on the full complement of chemoreceptors, we sought to determine whether inter-receptor communication can be detected without the aid of the synthetic ligand. We tested whether *E. coli* mutants lacking an MCP class exhibit diminished chemotactic responses. MCP mutants were treated with the attractant serine, and their behavioural responses were assessed by motion analysis. Responses to 100  $\mu$ M serine were decreased 12–14% in mutants lacking Tar or both Tar and Tap, relative to those in wild-type cells (Fig. 4b). Cells lacking Trg exhibited no detectable



**Figure 3** Response of wild-type *E. coli* (AW405) to chemoattractants after pre-treatment with buffer or ligands 1–3 at a galactose residue concentration of 10  $\mu$ M. **a**, *Escherichia coli* were allowed to adapt fully to ligands 1–3 (120 s) at 30 °C after pre-treatment. Results represent the average angular velocity over the first 2 min after the addition of serine. Full adaptation to serine concentrations of 100 nM or greater required more than 2 min, a result consistent with the adaptation time of *E. coli* strain AW405 in response to 300 nM aspartate, determined by motion analysis<sup>24</sup>. Values ( $\pm$ s.e.) are the average of at least four experiments performed in triplicate. **b**, Schematic model for the stabilization of MCP clusters and potentiation of serine responses. After treatment with multivalent ligand 3 (bottom), clusters of MCPs are stabilized relative to MCPs treated with monovalent ligand 1 (top). Trg (green) and GGBP (green) are shown. Additional MCPs are represented by the pink (Tsr) and yellow (Tar) receptors. Serine (red) is shown bound to Tsr. **c**, Motion analysis results after pre-treatment with ligand 3 at a final galactose residue concentration of 10  $\mu$ M. Values represent the average angular velocity over the first 2 min after addition of the second attractant (1  $\mu$ M serine, 1  $\mu$ M aspartate, or 10  $\mu$ M ribose). Adaptation times were at least 2 min in each case. Values ( $\pm$ s.e.) are the average of at least four experiments performed in triplicate.



**Figure 4** Response of *E. coli* to serine. **a**, Response of *E. coli* mutants to serine after pre-treatment with multivalent ligand 3 or buffer. Cells of *E. coli* strain AW405 (wild type, WT), RP2361 (*tar*<sup>-</sup>), RP5854 (*tar*<sup>-</sup> *tap*<sup>-</sup>), or AW550 (*ggbp*<sup>-</sup>) were pre-treated with either ligand 3 (10  $\mu$ M galactose residue concentration) or buffer. Serine was added to a final concentration of 1  $\mu$ M after adaptation to the galactose-substituted ligand (120 s). Results are the average of the angular velocity collected during the 2 min after addition of serine. The data were normalized to a control in which buffer was used for both treatments. Values ( $\pm$ s.e.) are the average of at least four experiments performed in triplicate. **b**, Response of *E. coli* MCP mutants to serine without pre-treatment. Serine (100  $\mu$ M) was added to cells of strain AW405 (WT), AW518 (*tsr*<sup>-</sup>), AW701 (*trg*<sup>-</sup>), RP2361 (*tar*<sup>-</sup>), or RP5854 (*tar*<sup>-</sup> *tap*<sup>-</sup>). Motion analysis was performed for 2 min. Values ( $\pm$ s.e.) are the average of three experiments performed in triplicate.

decrease in serine responses, a result consistent with the expectation that low-abundance receptors make smaller contributions to signal amplification. Notably, high attractant concentrations were required to distinguish between wild-type and mutant responses (see Fig. 4a). This finding highlights the benefits of synthetic multivalent ligand 3, which is a highly sensitive probe of the role of chemoreceptor clustering. Together, our results suggest that inter-receptor interactions can mediate the signal amplification that is required for the dynamic range observed for chemotactic responses<sup>1,3</sup>.

Despite the difference in receptor copy number, chemotactic responses towards galactose (through Trg) and serine (through Tsr) occur over comparable dynamic ranges<sup>1,2</sup>. Our finding that high- and low-abundance receptors can communicate to amplify signals indicates that information transfer occurs between these receptor classes. These results provide a mechanism for achieving parity among chemotactic responses despite the differences in receptor copy number. Our data complement the finding that intra-dimer methylation governs the dependence of the low-abundance receptors on the high-abundance MCPs for proper adaptation<sup>16</sup>. Thus, inter-receptor communication may be responsible for both signal amplification during excitatory responses and receptor methylation during adaptation.

Receptors are often localized to particular regions of the cell surface. Yet the significance of such localization is often obscure. We have shown that multivalent ligands can reveal the role of chemoreceptor arrays in governing cellular responses. Models for amplification through receptor arrays have been proposed for other signalling processes<sup>27,28</sup>, including T-cell-receptor-mediated lymphocyte activation<sup>29,30</sup>. Our results indicate that synthetic multivalent ligands can serve as valuable probes of the mechanisms underlying signal amplification and integration. □

Received 22 October; accepted 3 December 2001.

- Mesibov, R., Ordal, G. W. & Adler, J. The range of attractant concentrations for bacterial chemotaxis and the threshold and size over this range. *J. Gen. Physiol.* **62**, 203–223 (1973).
- Adler, J., Hazelbauer, G. L. & Dahl, M. M. Chemotaxis towards sugars in *Escherichia coli*. *J. Bacteriol.* **115**, 824–847 (1973).
- Jasuja, R., Yu-Lin, Trentham, D. R. & Khan, S. Response tuning in bacterial chemotaxis. *Proc. Natl Acad. Sci. USA* **96**, 11346–11351 (1999).
- Duke, T. A. J. & Bray, D. Heightened sensitivity of a lattice of membrane receptors. *Proc. Natl Acad. Sci. USA* **96**, 10104–10108 (1999).
- Bray, D., Levin, M. D. & Morton-Firth, C. J. Receptor clustering as a cellular mechanism to control sensitivity. *Nature* **393**, 85–88 (1998).
- Bren, A. & Eisenbach, M. How signals are heard during bacterial chemotaxis: Protein–protein interactions in sensory signal perception. *J. Bacteriol.* **182**, 6865–6873 (2000).
- Barkai, N. & Leibler, S. Robustness in simple biochemical networks. *Nature* **387**, 913–917 (1997).
- Hazelbauer, G. L. & Engström, P. Multiple forms of methyl-accepting chemotaxis proteins distinguished by a factor in addition to multiple methylation. *J. Bacteriol.* **145**, 35–42 (1981).
- Slocum, M. K. & Parkinson, J. S. Genetics of methyl-accepting chemotaxis proteins in *Escherichia coli*: organization of the *tar* region. *J. Bacteriol.* **155**, 565–577 (1983).
- Gegner, J. A., Graham, D. R., Roth, A. F. & Dahlquist, F. W. Assembly of an MCP receptor, CheW, and kinase CheA complex in the bacterial chemotaxis signal transduction pathway. *Cell* **70**, 975–982 (1992).
- Liu, Y., Levit, M., Lurz, R., Surette, M. G. & Stock, J. B. Receptor-mediated protein kinase activation and the mechanism of transmembrane signaling in bacterial chemotaxis. *EMBO J.* **16**, 7231–7240 (1997).
- Schuster, S. C., Swanson, R. V., Alex, L. A., Bourret, R. B. & Simon, M. I. Assembly and function of a quaternary signal transduction complex by surface plasmon resonance. *Nature* **365**, 343–347 (1993).
- Ottemann, K. M., Xiao, W., Shin, Y.-K. & Koshland, D. E. Jr A piston model for transmembrane signaling of the aspartate receptor. *Science* **285**, 1751–1754 (1999).
- Maddock, J. R. & Shapiro, L. Polar location of the chemoreceptor complex in the *Escherichia coli* cell. *Science* **259**, 1717–1723 (1993).
- Gestwicki, J. E. *et al.* Evolutionary conservation of methyl-accepting chemotaxis protein location in *Bacteria* and *Archaea*. *J. Bacteriol.* **182**, 6499–6502 (2000).
- Barnakov, A. N., Barnakov, L. A. & Hazelbauer, G. L. Comparison *in vitro* of a high- and a low-abundance chemoreceptor of *Escherichia coli*: similar kinase activation but different methyl-accepting activities. *J. Bacteriol.* **180**, 6713–6718 (1998).
- Li, J., Li, G. & Weis, R. M. The serine chemoreceptor from *Escherichia coli* is methylated through an inter-dimer process. *Biochemistry* **36**, 11851–11857 (1997).
- Gardina, P. J. & Manson, M. D. Attractant signaling by an aspartate chemoreceptor dimer with a single cytoplasmic domain. *Science* **274**, 425–426 (1996).
- Tatsuno, I., Homma, M., Oosawa, K. & Kawagishi, I. Signaling by the *Escherichia coli* aspartate chemoreceptor Tar with a single cytoplasmic domain per dimer. *Science* **274**, 423–425 (1996).
- Yeh, J. I., Biemann, H.-P., Pandit, J., Koshland, D. E. & Kim, S.-H. The three-dimensional structure of the ligand-binding domain of a wild-type chemotaxis receptor. *J. Biol. Chem.* **268**, 9787–9792 (1993).

- Trnka, T. M. & Grubbs, R. H. The development of  $L_2X_2Ru = CHR$  olefin metathesis catalysts: an organometallic success story. *Acc. Chem. Res.* **34**, 18–29 (2001).
- Gestwicki, J. E., Strong, L. E. & Kiessling, L. L. Tuning chemotactic responses using synthetic multivalent ligands. *Chem. Biol.* **7**, 583–591 (2000).
- Gestwicki, J. E. *et al.* Designed potent multivalent chemoattractants for *Escherichia coli*. *Bioorg. Med. Chem.* **9**, 2387–2393 (2001).
- Sager, B. M., Sekelsky, J. J., Matsumura, P. & Adler, J. Use of a computer to assay motility in bacteria. *Anal. Biochem.* **173**, 271–277 (1988).
- Parkinson, J. S. & Houts, S. E. Isolation and behavior of *Escherichia coli* deletion mutants lacking chemotaxis functions. *J. Bacteriol.* **151**, 106–113 (1982).
- Hazelbauer, G. L., Mesibov, R. E. & Adler, J. *Escherichia coli* mutants defective in chemotaxis toward specific chemicals. *Proc. Natl Acad. Sci. USA* **64**, 1300–1307 (1969).
- Yin, C. C. & Lai, F. A. Intrinsic lattice formation by ryanodine receptor calcium-release channel. *Nature Cell Biol.* **2**, 669–671 (2000).
- Matsuuchi, L. & Gold, M. R. New views of BCR structure and organization. *Curr. Opin. Immunol.* **13**, 270–277 (2001).
- Delon, J. & Germain, R. N. Information transfer at the immunological synapse. *Curr. Biol.* **10**, R923–R933 (2000).
- Chan, C., George, A. J. T. & Stark, J. Cooperative enhancement of specificity in a lattice of T cell receptors. *Proc. Natl Acad. Sci. USA* **98**, 5758–5763 (2001).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

## Acknowledgements

We thank J. Adler for supplying *E. coli* strains AW405, AW550, AW518 and AW701, and for advice. We thank J. S. Parkinson for *E. coli* strains RP1078, RP5854 and RP2361, and for anti-MCP antibodies. This work was supported by the National Institutes of Health (NIH). J.E.G. acknowledges the NIH Biotechnology Training Grant for support.

## Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.L.K. (e-mail: [kiessling@chem.wisc.edu](mailto:kiessling@chem.wisc.edu)).

# Two-component regulator of *Enterococcus faecalis* cytolysin responds to quorum-sensing autoinduction

Wolfgang Haas, Brett D. Shepard & Michael S. Gilmore

Department of Microbiology and Immunology, and Department of Ophthalmology, University of Oklahoma Health Sciences Center, 975 NE 10th Street, Oklahoma City, Oklahoma 73104, USA

Bacteria of the genus *Enterococcus* are the main causes of highly antibiotic-resistant infections that are acquired in hospitals<sup>1,2</sup>. Many clinical isolates of *Enterococcus faecalis* produce an exotoxin called cytolysin that contributes to bacterial virulence<sup>3</sup>. In addition to its toxin activity, the cytolysin is bactericidal for nearly all Gram-positive organisms<sup>4</sup>. An understanding of conditions that regulate cytolysin expression has advanced little since its initial description<sup>5</sup>. Here we show that the products of two genes, *cylR1* and *cylR2*, which lack homologues of known function, work together to repress transcription of cytolysin genes. Derepression occurs at a specific cell density when one of the cytolysin subunits reaches an extracellular threshold concentration. These observations form the basis of a model for the autoinduction of the cytolysin by a quorum-sensing mechanism involving a two-component regulatory system.

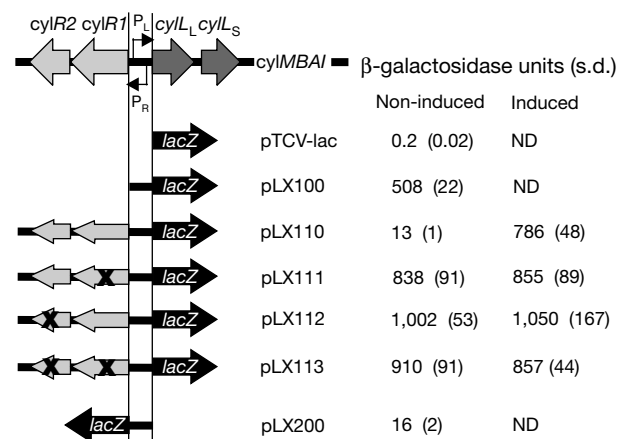
The large and small subunits of the cytolysin are encoded respectively by *cylL<sub>L</sub>* and *cylL<sub>S</sub>*, and are post-translationally modified by the *cylM* gene product. The modified cytolysin subunits are then



secreted from the cell by an ABC transporter encoded by *cylB*<sup>6</sup>. Extracellularly, the two subunits are activated by a protease encoded by *cylA*<sup>7</sup>, resulting in the fully functional mature cytolysin subunits *CylL*<sup>''</sup> and *CylS*<sup>''</sup>. An additional gene, *cylI*, encodes a protein that confers protection from the bactericidal activity to the cytolysin-producing cell<sup>8</sup>. Genes *cylL*, *cylS*, *cylM*, *cylB*, *cylA* and *cylI* are collinear, as shown in Fig. 1 (see also ref. 3).

*Enterococcus faecalis* cytolysin has been termed a 'pseudohaemolysin' as it produces an easily discernible zone of haemolysis on blood agar plates; however, little haemolytic activity could be detected in broth culture<sup>5</sup>. This suggests that expression of the cytolysin is regulated in a manner that is sensitive to specific environmental conditions. To determine the basis for regulation of the cytolysin operon we mapped the promoter initiating transcription by primer extension analysis (W.H., B.D.S. and M.S.G., manuscript in preparation), and we constructed *lacZ* reporter fusions<sup>9</sup> to allow a range of potentially inducing conditions to be screened. Fusion of the cytolysin operon promoter pL to the *lacZ* reporter gene resulted in a comparatively high level of *lacZ* expression by *E. faecalis* (plasmid pLX100 in Fig. 2). Two newly identified genes, *cylR1* and *cylR2*—natively occurring immediately upstream of the cytolysin promoter and oriented in the opposite direction of the cytolysin operon (GenBank accession number AF394225)—were included in the *lacZ* fusion pLX110. Expression from the pL promoter was nearly, but not completely, repressed in the presence of *cylR1* and *cylR2*. The products of both genes were required for repression, as frameshift mutations in either product resulted in derepression (plasmids pLX111–113; Fig. 2).

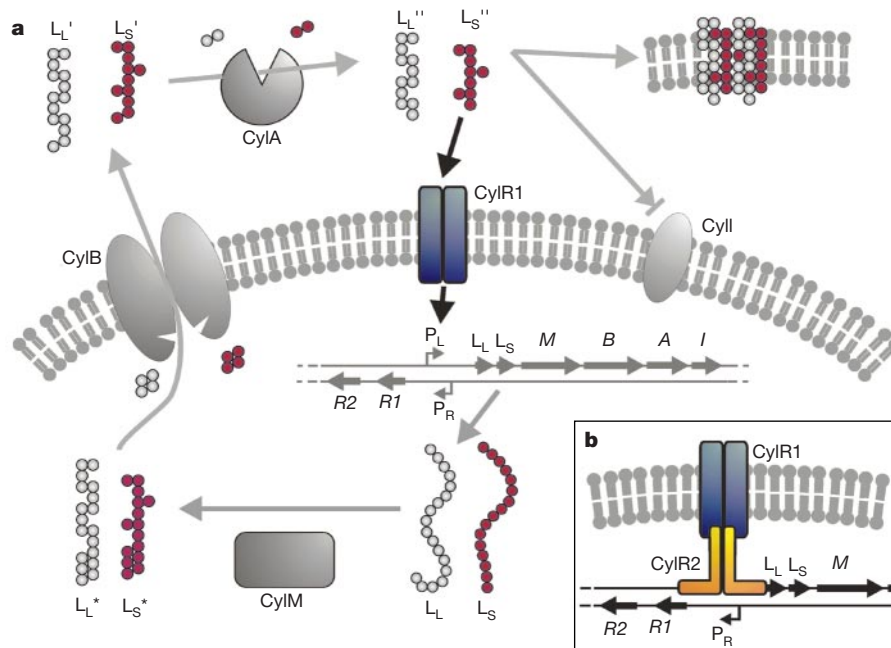
The *cylR1* gene encodes an inferred protein of 94 amino acids, and has no closely related homologues in GenBank. The protein is predicted to be  $\alpha$ -helical and to contain three transmembrane domains<sup>10–12</sup>. The deduced *CylR2* protein is 66 amino acids in size, is predicted to be a non-globular,  $\alpha$ -helical protein with a helix-



**Figure 2** Transcriptional fusions to the  $\beta$ -galactosidase (*lacZ*) reporter gene in the shuttle vector pTCV-lac<sup>9</sup>. DNA, including the cytolysin promoter and upstream regulatory genes, was cloned adjacent to a promoterless *lacZ* gene (pLX100 through to pLX113). Mutations in the open reading frames of *cylR1* and *cylR2* are indicated by an X. Promoter activity in the divergent direction was tested in pLX200. *lacZ* reporter expression was determined by a modified Miller assay, and expressed as optical density at 405 nm  $\times 10$  per mg protein. *CylS*<sup>''</sup> was added in the form of sterile supernatants from strain FA2-2 (pWH705) to induce samples, while culture supernatant from a negative control, strain FA2-2 (pAT28), harbouring only the cloning vector was added to the uninduced samples.

turn-helix DNA-binding motif<sup>3</sup>, and to belong to the helix-turn-helix family of proteins<sup>14,15</sup>.

Strain FA2-2 (pLX110), which possesses the cytolysin promoter pL together with the two regulatory genes fused to the  $\beta$ -galactosidase gene, was used as an indicator strain to identify environ-



**Figure 1** Model of cytolysin maturation and regulation. **a**, The peptide subunits *CylL* and *CylS* are ribosomally synthesized and modified post-translationally by *CylM*. The modified peptides, *CylL*<sup>\*</sup> and *CylS*<sup>\*</sup>, are proteolytically cleaved and secreted, followed by extracellular activation by a second protease, *CylA*. The two modified, trimmed and fully active peptide subunits *CylL*<sup>''</sup> and *CylS*<sup>''</sup> are both required for lysis of prokaryotic and eukaryotic cells<sup>7</sup>. The cytolysin-producing cell itself is protected by *CylI* by means of an unknown mechanism<sup>8</sup>. **b**, The *CylS*<sup>''</sup> subunit induces expression of the cytolysin operon, and *CylR1* and *CylR2* are both required for repression. In the simplest model, *CylS*<sup>''</sup> may interact with regulatory protein *CylR1* on the membrane surface, on the basis of its hydrophobic composition. *CylR1* may then function by enabling the predicted DNA-binding protein, *CylR2*, to repress transcription from the cytolysin promoter pL in an uninduced state. Although no other regulatory components are encoded by the cytolysin operon, we cannot exclude the possibility that other cellular factors are involved as intermediates in the signalling process.

mental factors that induce cytolyisin expression. Among the factors that we tested was a panel of *E. faecalis* strains variously expressing the cytolyisin or components thereof. A blue discoloration near strains producing cytolyisin demonstrated that only organisms producing cytolyisin were able to induce *lacZ* expression in the indicator strain (Fig. 3, squares A1 and A3). No *lacZ* expression was observed near filter disks saturated with non-cytolytic control strains (Fig. 3, squares A2 and B1). Strains that produce single toxin subunits CylL<sub>S</sub>' or CylL<sub>L</sub>', which lacked the terminal CylA-mediated proteolytic processing<sup>16</sup>, were unable to induce *lacZ* expression, demonstrating that the cytolyisin subunits need to be fully mature to achieve inducing activity (Fig. 3, squares B2 and B3). Strains that only produce CylL<sub>S</sub>' induced large zones of β-galactosidase expression in the indicator strain (Fig. 3, squares C1 and C2), whereas a strain expressing only mature CylL<sub>L</sub>' was incapable of induction (Fig. 3, square C3). Plasmid pWH705 contains a 4-base-pair (bp) deletion in the *cylL<sub>L</sub>* gene, resulting in a mutated primary CylL<sub>L</sub> translation product that shares the first 32 amino acids with the wild-type cytolyisin subunit, followed by 25 amino acids of missense information. To formally exclude any contribution of this residual portion of CylL<sub>L</sub>, we constructed a second mutated derivative of the operon (pWH617) that introduced a stop codon three codons in a 3' direction to the initiator. This construct was observed to be fully capable of induction of the cytolyisin promoter. These data demonstrate that the fully processed and activated cytolyisin subunit CylL<sub>S</sub>' (but not CylL<sub>S</sub>', CylL<sub>L</sub>' or CylL<sub>L</sub>') induces transcription from the pL promoter in the presence of the two regulatory proteins CylR1 and CylR2.

Addition of culture supernatant containing mature CylL<sub>S</sub>' to the indicator strain harbouring the cytolyisin promoter-regulated *lacZ* fusion resulted in an increase in β-galactosidase activity by 60-fold over the uninduced control (Fig. 2). Strains containing transcriptional fusions in which either or both *cylR1* and *cylR2* regulatory genes were mutationally inactivated, constitutively expressed high levels of *lacZ* in the absence of inducer, indicating that mutation in

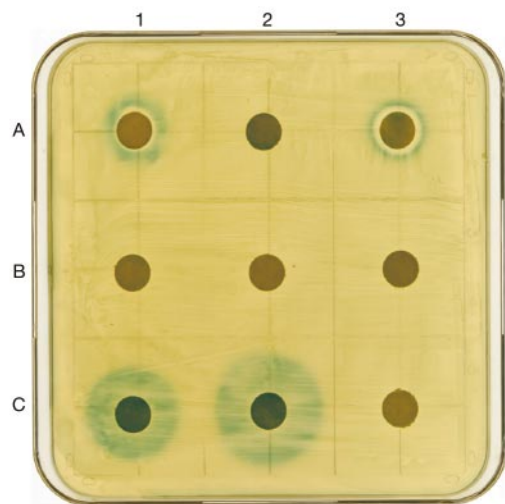
either of these regulatory genes results in complete derepression of the cytolyisin promoter.

Quorum-sensing systems function by producing low levels of an inducing substance that accumulates in the environment until a threshold is reached, at which point there is a change in cellular behaviour<sup>17</sup>. A sensitive bioassay—using the ability of CylL<sub>S</sub>' to induce β-galactosidase expression by an indicator strain that harbours a cytolyisin promoter fusion—was used to test for density-dependent changes in expression of the cytolyisin operon as it naturally occurs in *E. faecalis* FA2-2 (pAM714)<sup>18</sup>. Induction of the cytolyisin operon in planktonic brain heart infusion (BHI) culture occurred when the cell density exceeded 10<sup>7</sup> colony-forming units per ml (data not shown).

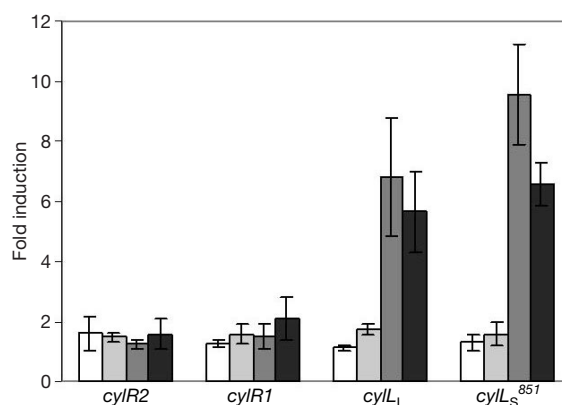
The effect of CylL<sub>S</sub>' on transcription of the cytolyisin genes was shown directly using strain FA2-2 (pWH851), in which the second codon of *cylL<sub>S</sub>* has been mutated to a stop codon and hence is incapable of CylL<sub>S</sub>' production or autoinduction. Sterile filtered culture supernatants containing various concentrations of wild-type CylL<sub>S</sub>' were added to growing cultures of the *cylL<sub>S</sub>*-deficient mutant, and total RNA was isolated 1 h after exposure and quantified by dot blot (not shown) and real-time polymerase chain reaction (PCR) analyses (Fig. 4). Transcription of the structural genes *cylL<sub>L</sub>* and *cylL<sub>S</sub>*<sup>851</sup> (the mutated form of *cylL<sub>S</sub>* harboured by strain FA2-2 (pWH851)) increased six–sevenfold after addition of mature CylL<sub>S</sub>' (Fig. 4). This increase in transcription was dependent on the amount of autoinducer added, with saturation being achieved by addition of as little as 2.5% v/v supernatant from a CylL<sub>S</sub>'-producing strain. Transcription of *cylR1* and *cylR2* from the divergent promoter was not detectably affected by exposure to CylL<sub>S</sub>'.

Quorum-sensing mechanisms, whereby a secreted autoinducer activates transcription at a specific cell density, have been described for a variety of organisms. The cytolyisin system, however, is unusual in that it does not seem to use a recognizable, highly conserved two-component signal transduction system consisting of a histidine kinase and a response regulator<sup>19</sup>. Instead, an apparently small helix-turn-helix DNA-binding protein, CylR2, and an apparent transmembrane protein of unknown function, CylR1, work together to repress the cytolyisin operon. Whether this regulation involves additional cellular intermediates in a signalling cascade remains to be determined. These results show that no additional genes occurring within the cytolyisin operon encode such intermediate signalling functions.

Two β-lactam resistance genes in *Staphylococcus* bacteria are regulated by transcriptional repressors that are proteolytically



**Figure 3** Induction of reporter gene expression by CylL<sub>S</sub>'. A lawn of the indicator strain, *E. faecalis* FA2-2 (pLX110), was plated out on X-gal containing BHI agar. *Enterococcus faecalis* FA2-2 strains containing various plasmids were grown in BHI broth for 6 h, impregnated onto Whatman paper disks, and placed on the agar plate. A blue coloration near the paper disk indicates induction of the β-galactosidase reporter gene in the surrounding indicator strain. A1: FA2-2 (pAM714), secretes wild-type cytolyisin; A2: FA2-2 (pAM771), isogenic, non-cytolytic control for pAM714; A3: FA2-2 (pCGC), cytolyisin operon cloned into shuttle vector pAT28; B1: FA2-2 (pAT28); B2: FA2-2 (pX9QATΔA), secretes pre-activated CylL<sub>S</sub>'; B3: FA2-2 (pX9QPATΔA), secretes pre-activated CylL<sub>L</sub>'; C1: FA2-2 (pWH617), secretes activated CylL<sub>S</sub>'; C2: FA2-2 (pWH705), secretes activated CylL<sub>S</sub>'; C3: FA2-2 (pWH851), produces activated CylL<sub>L</sub>'.



**Figure 4** CylL<sub>S</sub>' induces transcription of *cylL* and *cylL<sup>851</sup>*, but not *cylR1* or *cylR2*. The non-autoinducing strain *E. faecalis* FA2-2 (pWH851) was grown for 1 h in the presence of control supernatant (white bars), 0.25% (light grey bars), 2.5% (dark grey bars), or 25% (black bars) CylL<sub>S</sub>'-containing culture supernatant. RNA was isolated, normalized and quantified by real-time PCR and dot blot analyses.

inactivated in the presence of penicillin<sup>20</sup>. Although it is possible that the regulatory steps leading to transcription of cytolysin also involve a proteolytic step, no enzymatic activity of any kind has been demonstrated or predicted for the putative transmembrane protein CylR1.

The finding of a cell-density-dependent system regulating cytolysin expression may help to answer a long-standing question. It has been observed that cytolysin can be encoded on pheromone-responsive transmissible plasmids<sup>21</sup>. It is unclear how the cytolysin plasmid could be transferred efficiently to a recipient, which without the cytolysin operon is sensitive to the bactericidal effect. Pheromones that induce mating between donor and recipient cells are active at concentrations as low as  $5 \times 10^{-11}$  M (ref. 17), which may make possible the conjugation of plasmids to potential recipients before the cytolysin is induced to bactericidal levels at higher cell densities.

The finding that cytolysin expression is dependent on a quorum-sensing mechanism provides insight into new strategies that might be exploited in targeting a trait that has been shown to contribute to the course and severity of an infection for which there are few therapeutic choices<sup>22</sup>. □

## Methods

### Bacterial strains and growth conditions

*Enterococcus faecalis* strains FA2-2 and OG1 were grown in BHI supplemented with 500  $\mu\text{g ml}^{-1}$  spectinomycin, 1,000  $\mu\text{g ml}^{-1}$  kanamycin, or 50  $\mu\text{g ml}^{-1}$  erythromycin, depending on the presence of various plasmid-borne resistances.

### Cloning and DNA manipulations

Plasmid pXMA9AT (ref. 16) contained most of the cytolysin gene cluster except for the recently described *cyl* gene<sup>8</sup>. The missing fragment was amplified by PCR and ligated into the pXMA9AT vector. The resulting plasmid, containing the entire cytolysin gene cluster, was termed pCGC. The QuikChange site-directed mutagenesis kit (Stratagene), using pCGC as a template, was used to introduce premature stop codons into *cylL* (plasmid pWH617) or *cylS* (plasmid pWH851). Plasmid pWH705 was generated by *P*<sub>in</sub>AI digestion of pCGC, treatment with mung bean nuclease (New England Biolabs) and re-ligation, which resulted in a frameshift mutation in *cylL*.

Transcriptional fusions to a promoterless *lacZ* gene were made in the shuttle vector pTCV-*lac*<sup>9</sup>. DNA from the cytolysin operon was amplified by PCR and cloned upstream of the reporter gene. We generated a mutation in *cylR1* (pLX111) by PCR amplification as follows: a primer pair was designed to amplify *cylR2* and part of *cylR1* in pLX110, while two additional primers amplified the remainder of *cylR1*, the pL promoter and part of *lacZ*. The two PCR products, each containing part of *cylR1* and adjacent DNA, were digested with *Spe*I, ligated and cloned into pTCV-*lac*. This procedure introduced not only a *Spe*I site near the start codon of *cylR1*, but also a frameshift mutation and a stop codon. We used the same approach to generate pLX112, which contained a mutation in *cylR2*. Plasmid pLX113 was the result of a spontaneous frameshift mutation in pLX111 that occurred in the *cylR2* gene. All constructs were sequenced by the DNA sequencing facility at the University of Oklahoma Health Sciences Center.

### Reporter gene assay

We measured  $\beta$ -galactosidase expression according to ref. 23, with modifications. *Enterococcus faecalis* FA2-2 cells containing transcriptional *lacZ* fusions were grown in 25 ml BHI, and after 13 h at 37 °C they were divided into two 10-ml aliquots. One aliquot was induced by adding 2.5 ml of sterile filtered culture supernatant from the *CylL*<sup>+</sup>-producing strain FA2-2 (pWH705). Sterile culture supernatant (2.5 ml) from the negative-control strain FA2-2 (pAT28) was added to the second, uninduced culture. Cells were incubated for 3 h, washed, resuspended in 1.8 ml Z buffer (60 mM Na<sub>2</sub>HPO<sub>4</sub>, 40 mM NaH<sub>2</sub>PO<sub>4</sub>, 10 mM KCl, 1 mM MgSO<sub>4</sub>, 50 mM  $\beta$ -mercaptoethanol)<sup>23</sup> and disrupted in a Mini-Beadbeater (Biospec Products). Five hundred microlitres of cleared cell lysate, diluted 1:10 or 1:50 in Z buffer when necessary, was incubated with 100  $\mu\text{l}$  O-nitrophenyl- $\beta$ -D-galactopyranoside (ONPG) (Amresco) substrate (4 mg ml<sup>-1</sup> ONPG in 0.1 M phosphate buffer, pH 7.0) for 30 min at room temperature. The optical density at 405 nm of a 200  $\mu\text{l}$  sample was read in a Thermomax microplate reader (Molecular Devices) after adding 250  $\mu\text{l}$  of 1 M Na<sub>2</sub>CO<sub>3</sub> to stop the reaction. The protein concentration of 10  $\mu\text{l}$  of cell lysate was determined with 200  $\mu\text{l}$  Coomassie plus protein assay reagent (Pierce). The optical density at 405 nm per mg protein was multiplied by ten and expressed as  $\beta$ -galactosidase units. Assays were repeated in four independent experiments.

### RNA quantification

A culture of *E. faecalis* FA2-2 (pWH851) was grown to an optical density at 562 nm of 0.5 and divided into twelve 30-ml aliquots. Ten, one, 0.1 or 0 ml of sterile filtered culture supernatant from the *CylL*<sup>+</sup>-producing strain FA2-2 (pWH617) were added to each aliquot and volume-adjusted to 40 ml with sterile culture supernatant from a control

strain FA2-2 (pAT28). RNA was isolated<sup>24</sup> after incubation for 1 h at 37 °C and quantified using real-time PCR. Amplification, detection and analysis of *cylL*<sup>25</sup>, *cylL*, *cylR1* and *cylR2* messenger RNA were performed using the ABI Prism 7700 Sequence Detection System (PE Biosystems). Primers for each gene were designed to produce amplicons of equivalent length (100 bp). In preliminary reactions, 10 ng of each RNA sample was used to generate complementary DNA from 23S ribosomal RNA as an internal control with the TaqMan Reverse Transcription Reagents kit (PE Biosystems). The amount of total RNA was normalized when necessary and subjected to reverse transcription with primers for each individual gene. Five microlitres from each reverse transcription reaction were used in independent PCR amplifications reactions with the SYBR Green PCR Master Mix kit (PE Biosystems). Contamination of RNA samples by residual genomic DNA was ascertained by performing a cDNA synthesis reaction without reverse transcriptase, followed by PCR amplification. The threshold cycle (C<sub>T</sub>) determined by the SDS was converted into fold-induction units for each individual gene. The data represent the mean and standard deviation calculated from three independent samples. A dot blot analysis was performed in duplicate using a probe for *cylL*<sub>LS</sub> as described<sup>23</sup>. 23S rRNA served as loading control.

A detailed description of materials and methods is available as Supplementary Information; see also <http://www.enterococcus.ouhsc.edu>.

Received 9 August; accepted 4 October 2001.

1. National Nosocomial Infections Surveillance (NNIS) system report, data summary from January 1992–April 2000, issued June 2000. *Am. J. Infect. Contr.* **28**, 429–448 (2000).
2. Mundy, L. M., Sahm, D. F. & Gilmore, M. Relationships between enterococcal virulence and antimicrobial resistance. *Clin. Microbiol. Rev.* **13**, 513–522 (2000).
3. Haas, W. & Gilmore, M. S. Molecular nature of a novel bacterial toxin: the cytolysin of *Enterococcus faecalis*. *Med. Microbiol. Immunol. (Berl)* **187**, 183–190 (1999).
4. Brock, T. D., Peacher, B. & Pierson, D. Survey of the bacteriocins of enterococci. *J. Bacteriol.* **86**, 702–707 (1963).
5. Todd, E. W. A comparative serological study of streptolysins derived from human and from animal infections, with notes on pneumococcal hemolysin, tetanolysin and staphylococcus toxin. *J. Pathol. Bacteriol.* **39**, 299–321 (1934).
6. Gilmore, M. S., Segarra, R. A. & Booth, M. C. An HlyB-type function is required for expression of the *Enterococcus faecalis* hemolysin/bacteriocin. *Infect. Immun.* **58**, 3914–3923 (1990).
7. Booth, M. C. et al. Structural analysis and proteolytic activation of *Enterococcus faecalis* cytolysin, a novel antibiotic. *Mol. Microbiol.* **21**, 1175–1184 (1996).
8. Coburn, P. S., Hancock, L. E., Booth, M. C. & Gilmore, M. S. A novel means of self-protection, unrelated to toxin activation, confers immunity to the bactericidal effects of the *Enterococcus faecalis* cytolysin. *Infect. Immun.* **67**, 3339–3347 (1999).
9. Poyart, C. & Trieu-Cuot, P. A broad-host-range mobilizable shuttle vector for the construction of transcriptional fusions to  $\beta$ -galactosidase in gram-positive bacteria. *FEMS Microbiol. Lett.* **156**, 193–198 (1997).
10. Cserzo, M., Wallin, E., Simon, I., von Heijne, G. & Elofsson, A. Prediction of transmembrane  $\alpha$ -helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Eng.* **10**, 673–676 (1997).
11. Hofmann, K. & Stoffel, W. TMbase—a database of membrane spanning protein segments. *Biol. Chem. Hoppe-Seyler* **374**, 166 (1993).
12. Tusnady, G. E. & Simon, I. Principles governing amino acid composition of integral membrane proteins: application to topology prediction. *J. Mol. Biol.* **283**, 489–506 (1998).
13. Brennan, R. G. & Matthews, B. W. The helix-turn-helix DNA binding motif. *J. Biol. Chem.* **264**, 1903–1906 (1989).
14. Schultz, J., Milpetz, F., Bork, P. & Ponting, C. P. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl Acad. Sci. USA* **95**, 5857–5864 (1998).
15. Bateman, A. et al. The Pfam protein families database. *Nucleic Acids Res.* **28**, 263–266 (2000).
16. Gilmore, M. S. et al. Genetic structure of the *Enterococcus faecalis* plasmid pAD1-encoded cytolytic toxin system and its relationship to lantibiotic determinants. *J. Bacteriol.* **176**, 7335–7344 (1994).
17. Dunny, G. M. & Winans, S. C. *Cell–Cell Signaling in Bacteria* (ASM, Washington DC, 1999).
18. Ike, Y. & Clewell, D. B. Genetic analysis of the pAD1 pheromone response in *Streptococcus faecalis*, using transposon Tn917 as an insertional mutagen. *J. Bacteriol.* **158**, 777–783 (1984).
19. Hoch, J. A. Two-component and phosphorelay signal transduction. *Curr. Opin. Microbiol.* **3**, 165–170 (2000).
20. Zhang, H. Z., Hackbarth, C. J., Chansky, K. M. & Chambers, H. F. A proteolytic transmembrane signaling pathway and resistance to  $\beta$ -lactams in staphylococci. *Science* **291**, 1962–1965 (2001).
21. Francia, M. V. et al. Completion of the nucleotide sequence of the *Enterococcus faecalis* conjugative virulence plasmid pAD1 and identification of a second transfer origin. *Plasmid* **46**, 117–127 (2001).
22. Dong, Y. H. et al. Quenching quorum-sensing-dependent bacterial infection by an N-acyl homoserine lactonase. *Nature* **411**, 813–817 (2001).
23. Sambrook, J. & Russell, D. W. *Molecular Cloning* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2001).
24. Shepard, B. D. & Gilmore, M. S. Identification of aerobically and anaerobically induced genes in *Enterococcus faecalis* by random arbitrarily primed PCR. *Appl. Environ. Microbiol.* **65**, 1470–1476 (1999).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

### Acknowledgements

We thank P. Coburn, L. Hancock, M. Engelbert, C. Cox, K. Hatter, D. Johnson, B. Jett and K. Gilmore for technical support and critical review of the manuscript. This work was supported by grants from the National Institutes of Health and Research to Prevent Blindness.

Correspondence and requests for materials should be addressed to M.S.G. (e-mail: michael-gilmore@ouhsc.edu).



# Identification of a host protein essential for assembly of immature HIV-1 capsids

Concepcion Zimmerman\*, Kevin C. Klein†, Patti K. Kiser†, Aalok R. Singh\*, Bonnie L. Firestein‡, Shannyn C. Riba† & Jaisri R. Lingappa†§

\* Department of Physiology, University of California at San Francisco, San Francisco, California 94143, USA

† Department of Pathobiology, University of Washington School of Public Health; and § Department of Medicine, Division of Allergy and Infectious Disease, University of Washington School of Medicine Seattle, Washington 98195, USA

‡ Department of Cell Biology and Neuroscience, Rutgers University, Piscataway, New Jersey 08854, USA

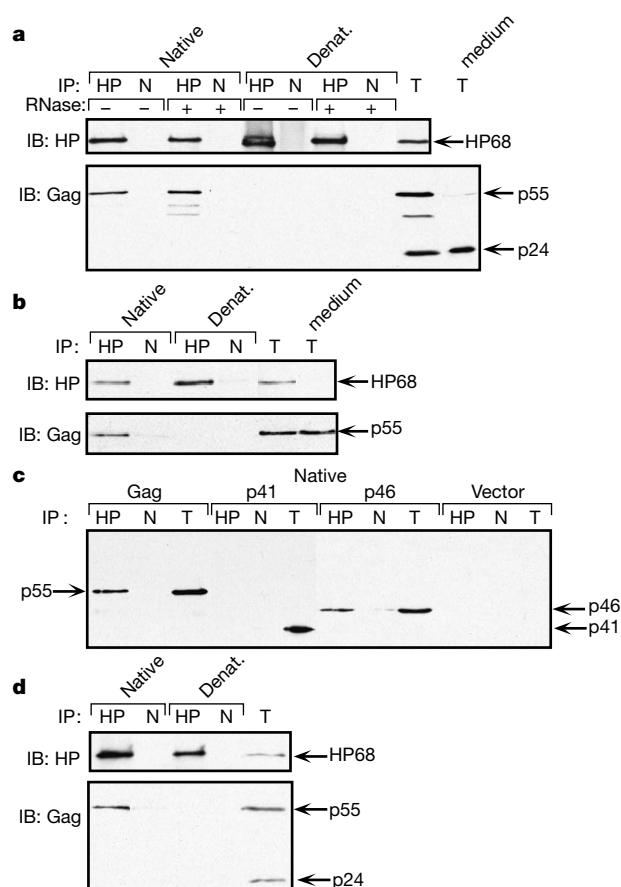
To form an immature HIV-1 capsid, 1,500 HIV-1 Gag (p55) polypeptides must assemble properly along the host cell plasma membrane. Insect cells and many higher eukaryotic cell types support efficient capsid assembly<sup>1</sup>, but yeast<sup>2</sup> and murine cells<sup>3,4</sup> do not, indicating that host machinery is required for immature HIV-1 capsid formation. Additionally, in a cell-free system that reconstitutes HIV-1 capsid formation, post-translational assembly events require ATP and a subcellular fraction<sup>5</sup>, suggesting a requirement for a cellular ATP-binding protein. Here we identify such a protein (HP68), described previously as an RNase L inhibitor<sup>6</sup>, and demonstrate that it associates post-translationally with HIV-1 Gag in a cell-free system and human T cells infected with HIV-1. Using a dominant negative mutant of HP68 in mammalian cells and depletion–reconstitution experiments in the cell-free system, we demonstrate that HP68 is essential for post-translational events in immature HIV-1 capsid assembly. Furthermore, in cells the HP68–Gag complex is associated with HIV-1 Vif, which is involved in virion morphogenesis and infectivity. These findings support a critical role for HP68 in post-translational events of HIV-1 assembly and reveal a previously unappreciated dimension of host–viral interaction.

Previously, we established a cell-free system for translation and assembly of HIV-1 Gag polypeptides into immature HIV-1 capsids—the protein shells that protect the viral genome<sup>5,7</sup>. In this system, ATP is required post-translationally for capsid assembly<sup>5</sup>. Because HIV-1 Gag is not thought to bind ATP, a cellular protein is probably involved. To determine whether a cellular nucleotide-binding protein is associated with assembling HIV-1 Gag chains, we examined whether antibodies directed against various ATP-binding molecular chaperones co-immunoprecipitate radiolabelled Gag in this system. Only one antibody (23c) co-immunoprecipitated Gag under native conditions, but not after denaturation (which disrupts protein–protein interactions). This antibody recognizes an epitope (containing LDD-COOH) present in several eukaryotic proteins, including the molecular chaperone TCP-1 (refs 8, 9). 23c failed to co-immunoprecipitate other substrates translated in the cell-free system, and recognized a 68K (relative molecular mass ( $M_r$ ) of 68,000) wheatgerm protein by immunoblotting and immunoprecipitation (see Supplementary Information Fig. 1).

Capsids formed during a cell-free assembly reaction closely correspond to authentic, immature HIV-1 capsids by biochemical and morphological criteria<sup>5</sup>. In the cell-free system, which uses wheatgerm extract as a source of cellular factors, 750S completed capsids are produced by means of a stepwise pathway of post-translational, Gag-containing complexes (10S, 80S, 150S and 500S). These complexes, which behave as true assembly intermediates<sup>5</sup>, are absent during the synthesis phase of the reaction; they accumulate after translation, and then diminish<sup>5,7</sup>. Co-immunoprecipitations,

using 23c, performed at different times during a pulse-chase assembly reaction suggested that Gag associates with HP68 only transiently, when assembly intermediates are at high levels, releasing HP68 once assembly is completed (see Supplementary Information Fig. 2). Co-immunoprecipitation of fractions produced by velocity sedimentation of assembly reaction products revealed that Gag was associated with HP68 in 80S and 500S (but not 10S or 750S) peak fractions (see Supplementary Information Fig. 2). Hepatitis B virus (HBV) core, which forms assembly intermediates and capsids that are fully assembled in the cell-free system<sup>10</sup>, was not co-immunoprecipitated by 23c (data not shown). Thus, HP68 associates selectively with partially assembled, post-translational HIV-1 capsid assembly intermediates, but not with unassembled Gag, fully assembled HIV-1 capsids, or HBV capsid assembly intermediates. These data suggest a specific role for HP68 in post-translational events of HIV-1 capsid assembly.

Wheatgerm HP68 (WGHP68) was isolated from wheatgerm extracts by immunoaffinity purification using 23c antibody, and microsequenced. Degenerate oligonucleotides were used to amplify a 2-kilobase (kb) complementary DNA from wheatgerm cDNA. The deduced amino-acid sequence (see Supplementary Information Fig. 3) was 71% identical to cDNA coding for the 68K human RNase L inhibitor<sup>6</sup> (termed here HuHP68). Both WGHP68 and HuHP68 contain two ATP/GTP-binding motifs<sup>11</sup> as well as the

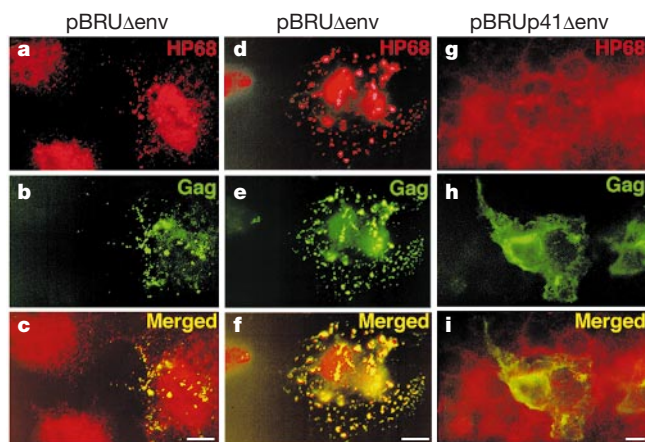


**Figure 1** Anti-HuHP68 co-immunoprecipitates HIV-1 Gag in mammalian cells. **a–d**, Native or denaturing (Denat.) immunoprecipitations (IP) on cell lysates using anti-HuHP68b (HP) or non-immune serum (N), followed by immunoblotting with antibody to HuHP68 (immunoblot (IB): HP) or Gag (IB: Gag). HIV-1 p24 and p55, 5% input cell lysate (T), and 10  $\mu$ l medium (T medium) are indicated. **a**, 293T cells transfected with pBRUΔenv, with or without RNase A treatment. **b**, Cos-1 cells expressing Gag. **c**, Cos-1 cells expressing Gag, an assembly-incompetent Gag mutant (p41), an assembly-competent Gag mutant (p46), or control vector (native immunoprecipitation only). **d**, ACH-2 cells chronically infected with HIV-1.

LDD-COOH epitope. HuHP68 binds and inhibits RNase L<sup>6</sup>—an interferon-dependent nuclease associated with polysomes<sup>12,13</sup> whose activation results in viral RNA degradation<sup>14</sup>. Antisense knockouts of RNase L inhibitor in cells that are infected with HIV-1 block virion production by reducing levels of HIV-1 RNA and HIV-1-specific proteins<sup>15</sup>. However, this fails to explain why WGHP68 binds to post-translational intermediates during cell-free HIV-1 capsid assembly. Thus, we investigated whether HuHP68 binds to and acts on fully synthesized Gag chains after translation in cells, in addition to binding and inhibiting RNase L as described previously<sup>6</sup>.

We generated polyclonal antisera that specifically recognize the carboxy terminus of WGHP68 or HuHP68 (see Supplementary Information Fig. 3). 293T cells were transfected with a plasmid (pBRUΔenv) encoding the entire HIV-1 genome except a region of the envelope gene<sup>16</sup>. Affinity-purified antisera to HuHP68 that was coupled to protein A beads (anti-HuHP68b) co-immunoprecipitated HIV-1 Gag from cell lysates under native conditions, but not after denaturation (Fig. 1a). Neither degradation of messenger RNA with RNase A (Fig. 1a) nor disassembly of ribosomes with EDTA (see below) affected HP68–Gag binding, indicating that this association does not require polysome integrity. Furthermore, HP68 did not associate with incomplete, nascent Gag chains (data not shown). Thus, HP68 appears to associate with Gag after translation. In addition, HP68–Gag binding occurred in the absence of other viral proteins, as seen in the cell-free system (see Supplementary Information Fig. 1) and in cells that express p55 Gag and produce immature HIV-1 virions<sup>17</sup> (Fig. 1b). Furthermore, HP68 bound an assembly-competent Gag mutant (p46) that lacks the p6 domain, but did not bind the assembly-incompetent p41 Gag mutant that lacks both the nucleocapsid (NC) and p6 domains<sup>5,18–21</sup> (Fig. 1c). Finally, HP68 associated with Gag in human T cells (ACH-2) that were producing infectious HIV-1 (Fig. 1d). Thus, HP68–Gag binding was seen in all of the HIV-1 capsid-producing systems that we examined.

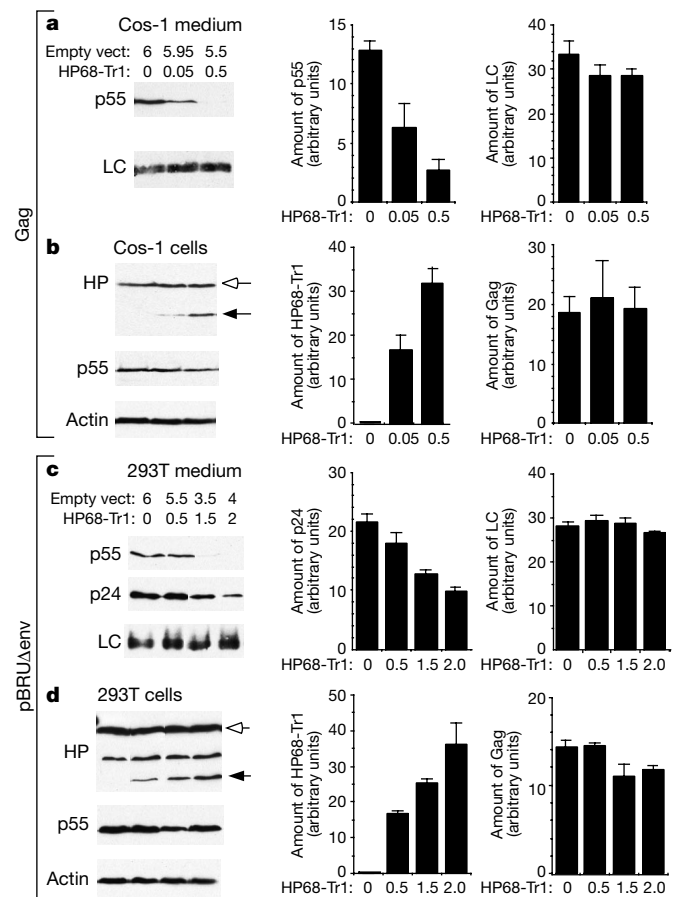
We confirmed HP68–Gag co-association using immunofluorescence microscopy. In Cos-1 cells, Gag was expressed in a predominantly clustered pattern (Fig. 2b, e), probably representing Gag localization to sites of virus formation at the plasma membrane. HP68 labelling revealed a diffuse pattern in all cells not expressing Gag (see Fig. 2a, two cells on the left), and a coarsely clustered pattern in Gag-expressing cells (Fig. 2d). Merged images revealed co-localization of HP68 with Gag in these clusters (Fig. 2c, f, yellow). This recruitment of HP68 by Gag was seen in 100% of



**Figure 2** Co-localization of HP68 with HIV-1 Gag in mammalian cells. Cos-1 cells were transfected with pBRUΔenv or pBRUp41Δenv (truncated proximal to the nucleocapsid domain in Gag), and double-label indirect immunofluorescence was performed. Cells were labelled for HP68 (a, d, g; red), or Gag (b, e, h; green). Images were merged to show overlap of HP68 and Gag labelling (c, f, i; yellow). Scale bars, 50 μm.

cells displaying HIV-1 Gag clusters, but not in any cells expressing the p41 assembly-incompetent Gag mutant (Fig. 2g–i).

To examine the function of HP68, we co-transfected truncated WGHP68 that encoded a stop codon before the second nucleotide-binding domain (WGHP68-Tr1) along with Gag expression plasmids into Cos-1 cells. We chose WGHP68 because it can be distinguished from endogenous HP68 using specific antisera. Furthermore, at low levels of expression it is less effective as an RNase L inhibitor than HuHP68 (data not shown), allowing post-transcriptional and post-translational effects of HP68 to be dissociated. Increasing expression of WGHP68-Tr1 resulted in a 4.7-fold dose-dependent decrease in the amount of HIV-1 Gag in the medium (Fig. 3a). Gag and actin levels in cell lysates remained unchanged (Fig. 3b), indicating that the effect of WGHP68-Tr1 is not mediated by changes in Gag synthesis or degradation, and that WGHP68-Tr1 is not toxic to cells. Similar results were obtained when pBRUΔenv and WGHP68-Tr1 were co-expressed in human cells (Fig. 3c, d). The dose-dependent decrease in p24 release was reversed on co-expression of wild-type WGHP68 with these plasmids (data not shown). The reduction in virion formation resulting from WGHP68-Tr1 expression while Gag levels are unchanged suggests that HP68 promotes virion formation by a post-translational mechanism. Co-immunoprecipitation of Gag with epitope-tagged WGHP68-Tr1 confirmed that WGHP68-Tr1 competes with wild-type HP68 for binding to



**Figure 3** Truncated HP68 blocks virion production. a–d, Cos-1 (a, b) or 293T (c, d) cells co-transfected with varying amounts (μg) of plasmid expressing WGHP68-Tr1 and empty vector, as indicated, plus plasmids for expression of HIV-1 Gag (a, b) or pBRUΔenv (c, d). Medium (a, c) was immunoblotted with Gag antibody (p55, p24), and re-probed with antibody to light-chain tracer (LC). Cell lysates (b, d) were immunoblotted using WGHP68 antiserum (HP) or Gag antibody (p55, p24), and re-probed using actin antibody (actin). White arrows indicate endogenous HP68; black arrows indicate WGHP68-Tr1 (a cross-reacting band can also be seen). Graphs represent blots from three experiments quantified using sample dilution standard curves.

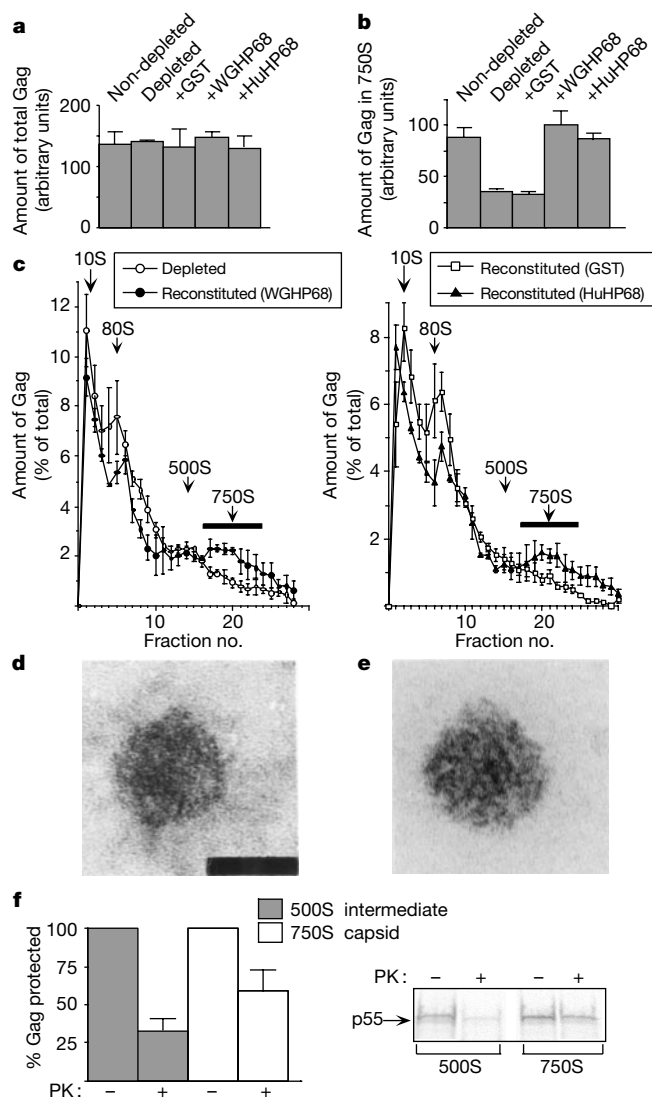
HIV-1 Gag (data not shown).

The dominant negative effect of WGHP68-Tr1 on virion formation (Fig. 3) indicates an essential role for HP68 in post-translational events in virion formation, but does not distinguish between HP68 acting on capsid assembly versus virus budding. For this, we used biochemical manipulation of the cell-free system. As wheat-germ extract does not contain functional RNase L<sup>6</sup>, HP68 in this system does not act as an inhibitor of RNase L. Immunodepletion of wheat-germ extract reduced levels of WGHP68 by 70% (data not shown). Immunodepletion had no effect on Gag translation (Fig. 4a), but reduced the amount of 750S completed capsids by 60% (Fig. 4b). Furthermore, the depleted reaction was arrested at the 500S assembly intermediate, with accumulation of 10S and 80S intermediates, but no formation of 750S completed capsids (Fig. 4b, c). The small amount of Gag seen in the 750S position using

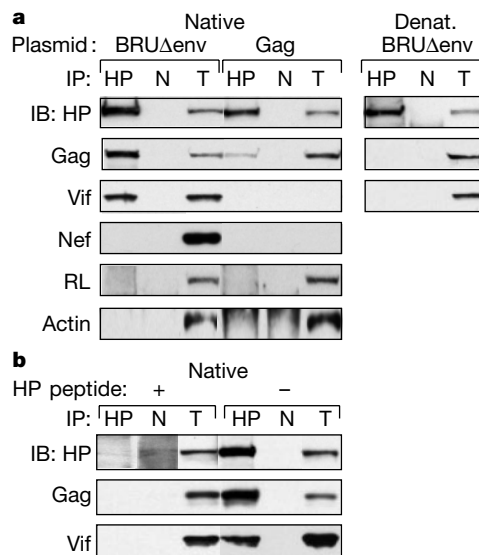
depleted extract was due to a trail from the 500S complex that enters the 750S region (Fig. 4b, c). Immunodepletion of HP68 had no effect on HBV capsid assembly (data not shown). Addition of purified recombinant WGHP68–glutathione *S*-transferase (GST) or HuHP68–GST to reactions programmed with immunodepleted extract produced a threefold increase in the amount of 750S capsid, restoring 750S capsids to levels seen in non-depleted extracts (Fig. 4c and data not shown). This reconstitution had no effect on Gag synthesis, indicating a post-translational action (Fig. 4a). Electron microscopy confirmed the absence of capsids in immunodepleted reactions (data not shown) and their presence in reconstituted reactions (Fig. 4d, e). Proteinase K digestion (Fig. 4f) revealed that the 500S capsid assembly intermediate (the end product of immunodepleted reactions) was protease-sensitive, whereas the 750S capsid produced by standard or reconstituted reactions was relatively resistant to protease treatment. This suggests that HP68-mediated conversion of Gag into completed 750S immature capsids is associated with a change in the conformation of the assembling capsid complex.

If HP68 acts post-translationally on capsid assembly intermediates, then other HIV-1 proteins involved in virion morphogenesis may also be associated with HP68. To test this, lysates of cells expressing pBRUΔenv were immunoprecipitated using anti-HuHP68b (in the presence and absence of EDTA) and immunoblotted for specific viral and cellular proteins. In addition to associating with Gag, HuHP68 associated with HIV-1 Gag-Pol and Vif, but not with Nef, which is probably incorporated into virions by direct association with the plasma membrane<sup>22–24</sup> (Fig. 5a and data not shown). Furthermore, neither actin nor RNase L were co-immunoprecipitated using anti-HuHP68b (Fig. 5a). Co-immunoprecipitation of Gag-Pol and Vif was inhibited by pre-incubating anti-HuHP68b with HuHP68 peptide, indicating the specificity of these interactions (Fig. 5b).

Our findings support a model in which HIV-1 upregulates a multifunctional cellular protein (HP68), exploiting it to promote critical events at two different stages in the viral life cycle. In



**Figure 4** HP68 depletion–reconstitution. **a**, **b**, Graphs showing total Gag synthesized (**a**) or amount of Gag in 750S completed capsids (**b**) from cell-free reactions programmed with indicated wheat-germ extracts. Non-depleted; depleted, immunodepleted of HP68; +GST, immunodepleted and reconstituted with GST alone; +WGHP68, depleted and reconstituted with WGHP68–GST; +HuHP68, depleted and reconstituted with HuHP68–GST. **c**, Amount of Gag in fractions from cell-free reactions in **a** that were subjected to velocity sedimentation. **d**, **e**, Transmission electron microscopy of capsids from immunodepleted cell-free reactions reconstituted with WGHP68–GST (**d**) or immature capsids from transfected mammalian cells (**e**). Scale bar, 100 nm. **f**, Proteinase K (PK) digestion of 500S and 750S fractions shown as percentage of Gag protected relative to normalized controls.



**Figure 5** Anti-HuHP68 co-immunoprecipitates HIV-1 Vif but not Nef or RNase L. **a**, Cos-1 cells transfected with pBRUΔenv or Gag plasmids were immunoprecipitated (IP) under native or denaturing (Denat.) conditions using anti-HuHP68b (HP) or non-immune serum (N), and immunoblotted (IB) with antibody to HuHP68 (HP), Gag, Vif, Nef, RNase L (RL), or actin. T, Total (5% of input cell lysate used in immunoprecipitation (HP, 10%)). The top of some actin lanes contain heavy chain cross-reacting to secondary antibody. **b**, Lysates of pBRUΔenv-transfected Cos-1 cells collected in 10 mM EDTA-containing buffer, and co-immunoprecipitated using beads pre-incubated with HuHP68 peptide or diluent control.



addition to acting post-transcriptionally by inhibiting RNase L<sup>6,15</sup>, HP68 binds Gag after translation and promotes its assembly into immature capsids. This is the first report, to our knowledge, of a cellular factor being essential for post-translational events in HIV-1 capsid assembly. Evidence for this post-translational mechanism comes from genetic (dominant negative), biochemical (depletion–reconstitution) and morphological (electron microscopy) studies presented here. HP68 appears to act during conversion of assembly intermediates into completed immature capsids, promoting a conformational change that may be important for immature capsid integrity or maturation. Thus, HP68 may act as a molecular chaperone during capsid assembly. Data demonstrating that HP68 selectively associates with three viral proteins critical for assembly of a fully infectious virion (Gag, Gag-Pol and Vif) provide further support for a post-translational action. These findings may help with understanding species-specific restriction to virion formation and the function of Vif, which acts during virion assembly by means of an unidentified host factor and is required for formation of virions that are fully infectious *in vivo*<sup>25–28</sup>. □

## Methods

### Cell-free assembly reactions

Cell-free translations, pulse chase and velocity sedimentation were performed as described<sup>5,7,10</sup>.

### Purification and sequencing of HP68

For purification, 3 ml of wheatgerm extract supernatant (centrifuged in a Beckman TL100.2 rotor, 15 min, 436,000g), was immunoprecipitated using 50 µg 23c antibody (Stressgen). Immunoprecipitate was transferred to polyvinylidene difluoride membrane; Coomassie staining revealed a single 23c-reactive protein (68K). This protein was microsequenced (ProSeq), found to be amino-terminally blocked, and was treated with cyanogen bromide and O-phthalaldehyde, which allowed Edman sequencing of two proline-containing peptides.

### cDNA amplification

Degenerate oligonucleotides corresponding to WGHP68 3' sequence were synthesized: 5'-ATGAATTC(ACGT)GG(CT)CT(GA)TA(CTG)GT(CTG)GG(GA)TC-3' and 5'-ATGAATTC(ACGT)GG(CT)CT(GA)TA(CTG)GT(CTG)GG(GA)TC-3'. WGHP68 was amplified by PCR, using wheatgerm cDNA (Invitrogen), 3' oligonucleotide corresponding to the WGHP68 C terminus, and 5' oligonucleotide corresponding to vector. This was performed multiple times, yielding a single 1.8-kb product that was ligated by TA cloning (Invitrogen). Non-coding ends were obtained through nested rapid amplification of cDNA ends (RACE) PCR. From overlapping cDNA clones, a complete sequence was obtained. The start was identified by a Kozak consensus sequence at the initiating methionine with two in-frame stop codons and no ATG codons upstream<sup>29</sup>, and by homology to the human homologue in GenBank<sup>6</sup>.

### Generation of antisera

Polyclonal rabbit antisera were generated against C-terminal peptides of human and wheatgerm HP68 (see Supplementary Information Fig. 3) and 19 N-terminal residues of human RNase L. HuHP68 antisera was affinity purified against HuHP68 C-terminal peptide and coupled to protein A, generating anti-HuHP68b.

### Transfections, immunoprecipitation and immunofluorescence

Cells in 6-cm dishes were transfected using Gibco lipofectamine (Cos-1) or lipofectamine plus (293T) with pCMVRev and PSVGagRRE-R (1 µg each) for HIV-1 Gag expression<sup>17</sup>, with pBRΔenv (2 µg, Fig. 3; 4 µg, Figs 1 and 5) for expression of the nearly complete HIV-1 genome<sup>16</sup>, or plasmids derived from these encoding indicated mutations. WGHP68-Tr1 was constructed by ligating cDNA encoding WGHP68 amino acids 1–378 into *NheI/XbaI* of pCDNA 3.1 (Invitrogen). Medium was collected 28 or 42 h after transfection for immunofluorescence and immunoblotting, respectively. For immunofluorescence, cells were fixed in paraformaldehyde, permeabilized with 1% triton, and incubated with mouse HIV-1 Gag antibody (1:50) and affinity-purified HuHP68 antiserum (1:2,000), followed by Cy3- and Cy2-coupled secondary antibody (Jackson) (1:200). We quantified 178 cells. In Fig. 3, rat immunoglobulin-γ was added to medium as a tracer (10 µg ml<sup>-1</sup>) at collection; immunoblots were performed with Gag antibody (Dako) at 1:500, WGHP68 antiserum at 1:500, or other antisera. Bands were quantified using an immunoblot standard curve generated with known sample dilutions. For immunoprecipitations (Figs 1 and 5), cells from 6-cm dishes were collected in 300 µl NP40 buffer<sup>5</sup> (containing no magnesium and 10 mM EDTA, Fig. 5), and 100 µl was immunoprecipitated with 50 µl anti-HuHP68b or beads coupled to non-immune serum. In Fig. 1a, indicated lysates were treated with RNase A (0.1 µg ml<sup>-1</sup>) at 22 °C for 15 min before immunoprecipitation. In Fig. 1d, ACH-2 cells<sup>30</sup> were stimulated with phorbol myristate acetate (10 nM) for 24 h before collection. Unstimulated ACH-2 cells, producing much lower levels of infectious virus, gave the same results (data not shown). In Fig. 5b,

anti-HuHP68b was pre-incubated with the HuHP68 peptide used to generate antisera and was then dissolved in dimethylsulphoxide (200 µM), or in dimethylsulphoxide alone.

### Immunodepletion–reconstitution

Wheatgerm extract (150 µl) was immunodepleted for 45 min at 4 °C with 100 µl beads coupled to antibody to WGHP68. Cell-free reactions (15 µl) were programmed using non-depleted wheatgerm or depleted wheatgerm. To some reactions containing depleted wheatgerm, WGHP68–GST, HuHP68–GST or GST alone was added (2 µl of approximately 20 ng µl<sup>-1</sup>) at the start of the reaction. After 3 h at 26 °C, 1% NP40 was added and reactions underwent velocity sedimentation (5 ml, 15–60% sucrose gradients, Beckman ML550 rotor, 45 min, 129,000g). Thirty fractions, collected using a fractionator, were analysed by SDS–polyacrylamide gel electrophoresis (PAGE) and autoradiography, followed by densitometry of Gag in each lane. Graphs show average of three independent experiments (± s.e.m.). For the protease digestions, aliquots of 500S and 750S peaks were incubated for 10 min at 22 °C with or without 0.1 µg ml<sup>-1</sup> proteinase K, and analysed by SDS–PAGE and autoradiography (quantification as above).

WGHP68 and HuHP68 were subcloned into pGEX (Pharmacia) encoding N-terminal GST fusion proteins. Expression was induced with 1 mM isopropyl β-D-thiogalactopyranoside; sarcosyl (0.5%) and phenylmethylsulphonyl fluoride (PMSF) (0.75 mM) were added after sonication. Supernatant (17,000 µg) was incubated with glutathione beads and eluted with 40 mM glutathione in 50 mM Tris, pH 8.0.

### Electron microscopy

Two cell-free reactions were programmed with HIV-1 Gag transcript and immunodepleted wheatgerm, and WGHP68–GST was added to one of the reactions. Cell-free reactions and medium from Cos-1 cells were transfected to produce immature HIV-1 virus<sup>17</sup>, treated with 1% NP40 to remove envelopes, and centrifuged on 2 ml 20–66% sucrose gradients (Beckman TLS55 rotor, 35 min, 135,000g). Fractions containing 750S peaks<sup>5</sup> were placed on Formvar-coated grids previously incubated with Gag antibody for 10 min, fixed in half-strength Karnovsky's fixative, stained with uranyl acetate, and examined by transmission electron microscopy (Japan Electric Optics Laboratories 1010).

Received 24 May; accepted 24 September 2001.

- Boulanger, P. & Jones, I. Use of heterologous expression systems to study retroviral morphogenesis. *Curr. Top. Microbiol. Immunol.* **214**, 237–260 (1996).
- Jacobs, E., Gheysen, D., Thines, D., Francotte, M. & de Wilde, M. The HIV-1 Gag precursor Pr55gag synthesized in yeast is myristoylated and targeted to the plasma membrane. *Gene* **79**, 71–81 (1989).
- Mariani, R. *et al.* A block to human immunodeficiency virus type 1 assembly in murine cells. *J. Virol.* **74**, 3859–3870 (2000).
- Mariani, R. *et al.* Mouse-human heterokaryons support efficient human immunodeficiency virus type 1 assembly. *J. Virol.* **75**, 3141–3151 (2001).
- Lingappa, J. R., Hill, R. L., Wong, M. L. & Hegde, R. S. A multistep, ATP-dependent pathway for assembly of human immunodeficiency virus capsids in a cell-free system. *J. Cell Biol.* **136**, 567–581 (1997).
- Bisbal, C., Martinand, C., Silhol, M., Lebleu, B. & Salehzada, T. Cloning and characterization of a RNase L inhibitor. A new component of the interferon-regulated 2-5A pathway. *J. Biol. Chem.* **270**, 13308–13317 (1995).
- Singh, A. R., Hill, R. L. & Lingappa, J. R. Effect of mutations in Gag on assembly of immature human immunodeficiency virus type 1 capsids in a cell-free system. *Virology* **279**, 257–270 (2001).
- Hynes, G. *et al.* Analysis of chaperonin-containing TCP-1 subunits in the human keratinocyte two-dimensional protein database: further characterisation of antibodies to individual subunits. *Electrophoresis* **17**, 1720–1727 (1996).
- Willison, K. *et al.* The T complex polypeptide 1 (TCP-1) is associated with the cytoplasmic aspect of Golgi membranes. *Cell* **57**, 621–632 (1989).
- Lingappa, J. R. *et al.* A eukaryotic cytosolic chaperonin is associated with a high molecular weight intermediate in the assembly of hepatitis B virus capsid, a multimeric particle. *J. Cell Biol.* **125**, 99–111 (1994).
- Traut, T. W. The functions and consensus motifs of nine types of peptide segments that form different types of nucleotide-binding sites. *Eur. J. Biochem.* **222**, 9–19 (1994).
- Salehzada, T., Silhol, M., Lebleu, B. & Bisbal, C. Polyclonal antibodies against RNase L. Subcellular localization of this enzyme in mouse cells. *J. Biol. Chem.* **266**, 5808–5813 (1991).
- Zhou, A., Hassel, B. A. & Silverman, R. H. Expression cloning of 2-5A-dependent RNase: a uniquely regulated mediator of interferon action. *Cell* **72**, 753–765 (1993).
- Player, M. R. & Torrence, P. F. The 2-5A system: modulation of viral and cellular processes through acceleration of RNA degradation. *Pharmacol. Ther.* **78**, 55–113 (1998).
- Martinand, C. *et al.* RNase L inhibitor is induced during human immunodeficiency virus type 1 infection and down regulates the 2-5A/RNase L pathway in human T cells. *J. Virol.* **73**, 290–296 (1999).
- Kimpton, J. & Emerman, M. Detection of replication-competent and pseudotyped human immunodeficiency virus with a sensitive cell line on the basis of activation of an integrated β-galactosidase gene. *J. Virol.* **66**, 2232–2239 (1992).
- Smith, A. J., Srinivasakumar, N., Hammarskjöld, M. L. & Rekosh, D. Requirements for incorporation of Pr160gag-pol from human immunodeficiency virus type 1 into virus-like particles. *J. Virol.* **67**, 2266–2275 (1993).
- Gheysen, D. *et al.* Assembly and release of HIV-1 precursor Pr55gag virus-like particles from recombinant baculovirus-infected insect cells. *Cell* **59**, 103–112 (1989).
- Hockley, D. J., Nermut, M. V., Grief, C., Jowett, J. B. & Jones, I. M. Comparative morphology of Gag protein structures produced by mutants of the gag gene of human immunodeficiency virus type 1. *J. Gen. Virol.* **75**, 2985–2997 (1994).
- Jowett, J. B., Hockley, D. J., Nermut, M. V. & Jones, I. M. Distinct signals in human immunodeficiency virus type 1 Pr55 necessary for RNA binding and particle formation. *J. Gen. Virol.* **73**, 3079–3086 (1992).
- Royer, M. *et al.* Functional domains of HIV-1 gag-polyprotein expressed in baculovirus-infected cells. *Virology* **184**, 417–422 (1991).
- Bukovsky, A. A., Dorfman, T., Weimann, A. & Gottlinger, H. G. Nef association with human immunodeficiency virus type 1 virions and cleavage by the viral protease. *J. Virol.* **71**, 1013–1018 (1997).

23. Pandori, M. W. *et al.* Producer-cell modification of human immunodeficiency virus type 1: Nef is a virion protein. *J. Virol.* **70**, 4283–4290 (1996).
24. Welker, R., Harris, M., Cardel, B. & Krausslich, H. G. Virion incorporation of human immunodeficiency virus type 1 Nef is mediated by a bipartite membrane-targeting signal: analysis of its role in enhancement of viral infectivity. *J. Virol.* **72**, 8833–8840 (1998).
25. Cohen, E. A., Subbramanian, R. A. & Gottlinger, H. G. Role of auxiliary proteins in retroviral morphogenesis. *Curr. Top. Microbiol. Immunol.* **214**, 219–235 (1996).
26. Madani, N. & Kabat, D. An endogenous inhibitor of human immunodeficiency virus in human lymphocytes is overcome by the viral Vif protein. *J. Virol.* **72**, 10251–10255 (1998).
27. Simon, J. H. & Malim, M. H. The human immunodeficiency virus type 1 Vif protein modulates the postpenetration stability of viral nucleoprotein complexes. *J. Virol.* **70**, 5297–5305 (1996).
28. Simon, J. H. *et al.* The regulation of primate immunodeficiency virus infectivity by Vif is cell species restricted: a role for Vif in determining virus host range and cross-species transmission. *EMBO J.* **17**, 1259–1267 (1998).
29. Kozak, M. Interpreting cDNA sequences: some insights from studies on translation. *Mamm. Genome* **7**, 563–574 (1996).
30. Folks, T. M. *et al.* Tumor necrosis factor  $\alpha$  induces expression of human immunodeficiency virus in a chronically infected T-cell clone. *Proc. Natl Acad. Sci. USA* **86**, 2365–2368 (1989).

**Supplementary Information** accompanies the paper on Nature's website (<http://www.nature.com>).

## Acknowledgements

Vif monoclonal antibody (TG001; Transgene) and Nef monoclonal antibody (EH1; gift of J. Hoxie) were provided by the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, National Institutes of Health. We thank J. Kucinski and F. Calayag for technical assistance; M. Hayden, T. Liegler, R. Grant and the Gladstone Institute, San Francisco, for assistance with HIV-infected cells; V. Kewalramani, D. Littman, M. Goldsmith and W. Hansen for advice or reagents; L. Caldwell and the Electron Microscopy Laboratory at the Fred Hutchinson Cancer Research Center in Seattle; and V. Lingappa, J. Lingappa, J. Dooher, J. Overbaugh, J. M. McCune, M. Linial and R. Hegde for discussions. This work was supported by grants to J.R.L. from the National Institutes of Health AIDS Division, Pediatric AIDS Foundation, and the University of California Universitywide AIDS Research Program.

## Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.R.L. (e-mail: [jais@u.washington.edu](mailto:jais@u.washington.edu)). The sequence for WGHP68 has been deposited in GenBank under accession number AY059462.

# IRE1 couples endoplasmic reticulum load to secretory capacity by processing the *XBP-1* mRNA

Marcella Calton, Huiqing Zeng, Fumihiko Urano, Jeffery H. Till, Stevan R. Hubbard, Heather P. Harding, Scott G. Clark & David Ron

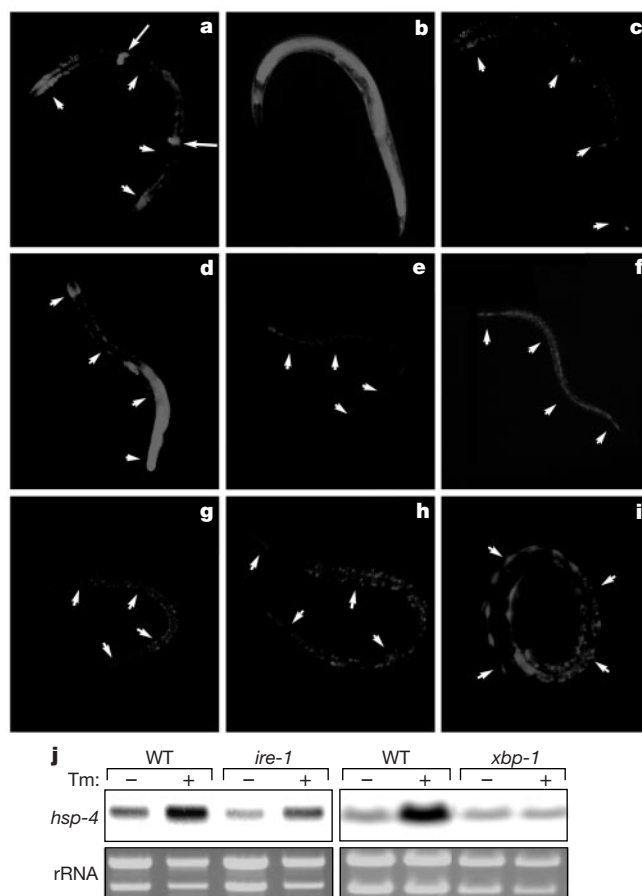
Skirball Institute of Biomolecular Medicine, New York University School of Medicine, New York, New York 10016, USA

The unfolded protein response (UPR), caused by stress, matches the folding capacity of endoplasmic reticulum (ER) to the load of client proteins in the organelle<sup>1,2</sup>. In yeast, processing of *HAC1* mRNA by activated Ire1 leads to synthesis of the transcription factor Hac1 and activation of the UPR<sup>3</sup>. The responses to activated IRE1 in metazoans are less well understood. Here we demonstrate that mutations in either *ire-1* or the transcription-factor-encoding *xbp-1* gene abolished the UPR in *Caenorhabditis elegans*. Mammalian *XBP-1* is essential for immunoglobulin secretion and development of plasma cells<sup>4</sup>, and high levels of *XBP-1* messenger RNA are found in specialized secretory cells<sup>5</sup>. Activation of the UPR causes IRE1-dependent splicing of a small intron from the *XBP-1* mRNA both in *C. elegans* and mice. The protein encoded by the processed murine *XBP-1* mRNA accumulated during the UPR, whereas the protein encoded by unprocessed mRNA did not. Purified mouse IRE1 accurately cleaved *XBP-1* mRNA *in vitro*,

indicating that *XBP-1* mRNA is a direct target of IRE1 endonucleolytic activity. Our findings suggest that physiological ER load regulates a developmental decision in higher eukaryotes.

IRE1 is a stress-activated endonuclease resident in the ER that is conserved in all known eukaryotes. In yeast, Ire1-mediated unconventional splicing of an intron from *HAC1* mRNA controls expression of the encoded transcription factor<sup>3</sup> and is required for upregulation of most UPR target genes<sup>6,7</sup>. UPR gene expression in mammals relies largely on pancreatic ER kinase (PERK) and ATF6 (refs 8–11), which are absent from yeast. Furthermore, mammalian IRE1 proteins activate Jun amino-terminal kinase by recruiting the TRAF2 protein to the ER membrane independently of their endonucleolytic activity<sup>12</sup>. It is unclear, therefore, whether IRE1 proteins of higher eukaryotes also signal ER stress through processing of *HAC1*-like mRNA targets. To address this issue we used a genetic strategy to identify UPR regulatory genes in *C. elegans*, a simple organism whose genome is predicted to encode homologues of all three known proximal, stress-sensing components of the metazoan UPR: *IRE1*, *PERK* and *ATF6*.

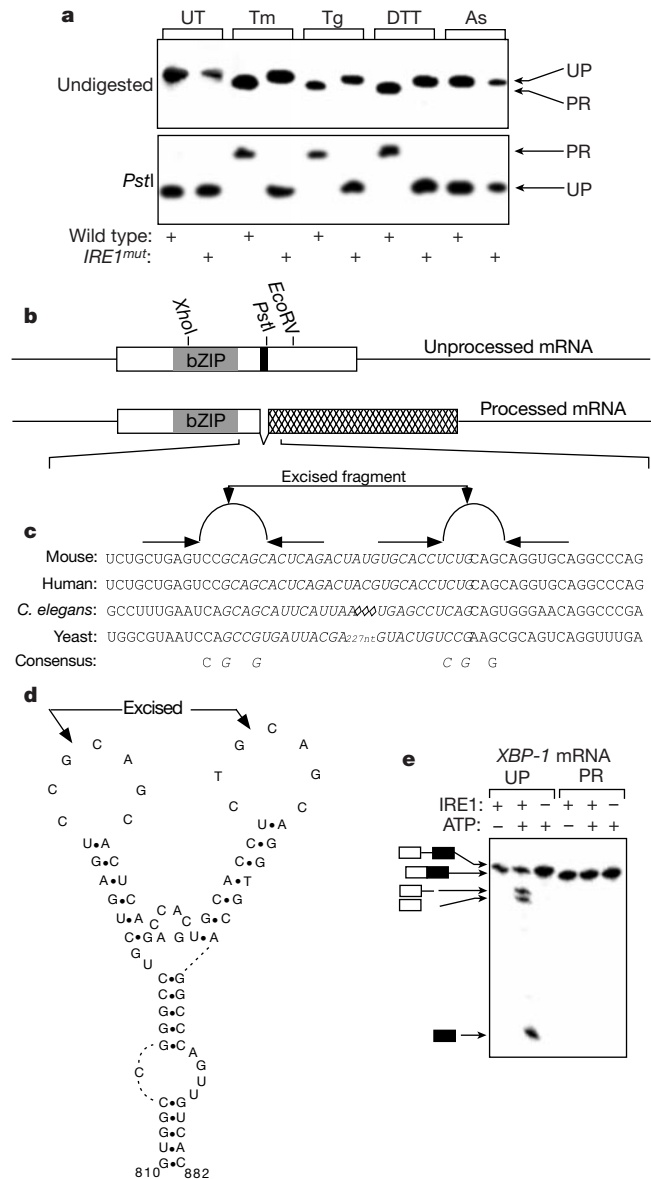
*Caenorhabditis elegans* has two homologues of the ER chaperone



**Figure 1** Identifying mutants in the *C. elegans* UPR. **a**, Fluorescent photomicrograph of an untreated adult *hsp-4::gfp(zcls4)V* transgenic animal. White arrowheads (in all panels) track the outline of the body; arrows indicate the spermatheca. **b**, An *hsp-4::gfp(zcls4)V* animal treated with tunicamycin. **c**, Tunicamycin-treated *ire-1(RNAi)III; hsp-4::gfp(zcls4)V* animal. **d**, Untreated *hsp-4::gfp(zcls4)V; upr-1(zc6)X* animal (note constitutive activation of the UPR in the posterior gut). **e**, Untreated *hsp-4::gfp(zcls4)V; upr-1(zc14)II* animal. **f**, Tunicamycin-treated *ire-1(zc14)III; hsp-4::gfp(zcls4)V* animal. **g**, Untreated *hsp-4::gfp(zcls4)V; upr-1(zc6)X* animal with a second mutation in *xbp-1(zc12)III*. **h**, Tunicamycin-treated *xbp-1(zc12)III; hsp-4::gfp(zcls4)V* animal. **i**, Tunicamycin-treated *xbp-1(RNAi)III; hsp-4::gfp(zcls4)V* animal. **j**, Northern blot of *hsp-4* RNA from untreated and tunicamycin-treated (Tm) wild-type (WT), *ire-1(zc10)II* or *xbp-1(zc12)III* mutant animals. Integrity and loading of the RNA is revealed by ethidium bromide staining of the ribosomal bands. rRNA, ribosomal RNA.

We found four *ise-1* mutations that impaired *hsp-4::gfp* induction (Fig. 1e, f, j); *zc10* and *zc11* alter the 3' splice acceptor site of intron 2, and *zc13* and *zc14* are missense mutations that affect conserved residues in the kinase domain (G702D and G739R, respectively). A fifth mutation, *zc12*, that strongly blocked *hsp-4::gfp* gene induction (Fig. 1g, h, j) mapped to the interval between cosmid F34D10 and *dpy-17* on chromosome III. A plausible candidate for a downstream

The predicted R74.3 protein of *C. elegans* is most similar to mouse XBP-1 (X-box binding protein-1), which binds the BiP ER



**Figure 2** Mammalian XBP-1 expression during ER stress is dependent on IRE1. **a**, XBP-1 immunoblot from murine fibroblasts treated with tunicamycin (Tm, upper panel) or thapsigargin (Tg, middle panel). CHOP immunoblot of the thapsigargin-treated lysates provides a positive control for the induction of the UPR (lower panel). **b**, XBP-1 immunoblot of lysates of differentiating mouse spleen cells cultured in medium with or without bacterial lipopolysaccharide (LPS) for the indicated number of days, and lysates of untreated and tunicamycin-treated fibroblasts. CREB (lower panel) serves as a loading and recovery control. MEF, mouse embryonic fibroblast. **c**, XBP-1, CHOP and CREB immunoblots and *XBP-1* and  $\beta$ -actin northern blots from untreated and tunicamycin-treated cells with the indicated genotypes.



stress-response element<sup>13</sup>. Therefore, we investigated further the regulation of XBP-1 in the mammalian UPR. ER stress induced a protein with a relative molecular mass of 54,000 ( $M_r$  54K) that was strongly reactive with antiserum to XBP-1 (Fig. 2a). The 54K XBP-1 protein was also induced during *in vitro* differentiation of B cells to plasma cells by exposure to bacterial lipopolysaccharide (Fig. 2b). Thus, ER stress- and differentiation-induced remodelling of the secretory apparatus are both associated with XBP-1 expression.

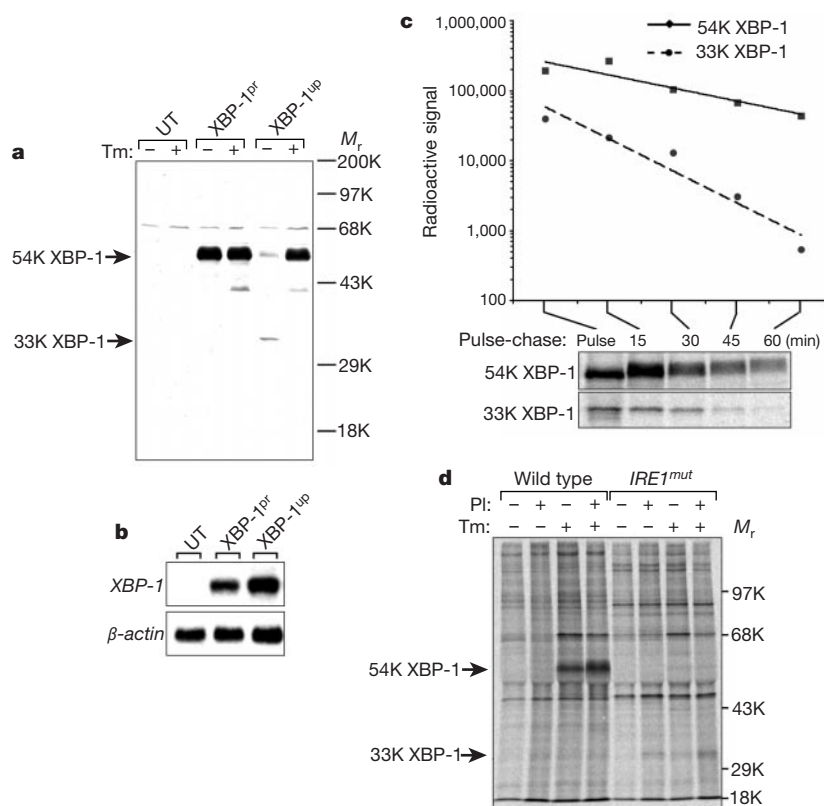
Mice have two *IRE1* genes: *IRE1 $\alpha$*  is essential for viability and is broadly expressed<sup>12,14</sup>, and *IRE1 $\beta$*  is expressed only in the gastrointestinal mucosa<sup>15,16</sup>. Induction of the 54K XBP-1 protein by ER stress was not observed in fibroblasts that lacked *IRE1* gene function (*IRE1 $\alpha$* <sup>-/-</sup> or *IRE1 $\alpha\beta$* <sup>-/-</sup>) but could be rescued by introduction of an *IRE1* transgene (Fig. 2c). ER stress has been shown to increase levels of *XBP-1* mRNA<sup>10</sup>. However, *IRE1* mutant cells that do not express the 54K XBP-1 protein had higher levels of *XBP-1* mRNA than wild-type cells (Fig. 2c, lower panels). Furthermore, *PERK*<sup>-/-</sup> cells that failed to induce *XBP-1* mRNA accumulate 54K XBP-1 protein when stressed (albeit to lower levels than wild-type cells, Fig. 2c). *IRE1* activity is thus required for expression of XBP-1, whereas ER-stress-mediated increase in *XBP-1* mRNA is less important in this process.

To investigate possible *IRE1*-mediated processing of *XBP-1* mRNA, we performed polymerase chain reaction with reverse transcription (RT-PCR) of *XBP-1* mRNA from unstressed and stressed wild-type and *IRE1*-mutant mouse cells. We found a consistent *IRE1* and ER stress-dependent decrease in the size of the *XhoI*–*EcoRV* fragment of the *XBP-1* complementary DNA (Fig. 3a). Sequencing revealed excision of a 26-nucleotide intron from the *XBP-1* cDNA derived from stressed wild-type cells (Fig. 3b, c). A similarly located 23-nucleotide intron was removed by ER stress from the *C. elegans xbp-1* mRNA (Fig. 3c, and data not

shown). The boundaries of these introns were encompassed in a predicted RNA structure that included two loops of seven residues held in place by short stems (Fig. 3c, d). The sequence of the loops corresponded perfectly to the empirically defined consensus sequence for cleavage of yeast *HAC1* mRNA by Ire1 (refs 17, 18). The RNA sequences 5' and 3' to these two putative processing sites are predicted to form extensive base-pair interactions that could hold together the cleaved ends of the mRNA (Fig. 3d), as has been predicted for the *HAC1* mRNA<sup>17</sup>.

To study the role of *IRE1* in the processing of *XBP-1* mRNA, we reconstructed the cleavage event *in vitro*. The cytoplasmic effector domain of murine *IRE1 $\beta$*  was purified from insect cells and reacted *in vitro* with a transcribed, unprocessed *XBP-1* mRNA that had been radiolabelled (*XBP-1*<sup>up</sup>), or with *XBP-1* mRNA transcribed from a processed cDNA template (*XBP-1*<sup>pr</sup>). *IRE1 $\beta$*  efficiently cleaved *XBP-1*<sup>up</sup> RNA in an ATP-dependent manner but did not cleave *XBP-1*<sup>pr</sup> (Fig. 3e). The location of the *in vitro* cleavage sites was mapped by primer extension and found to coincide with the sites used *in vivo* (Supplementary Information Fig. 1).

Re-ligation of *XBP-1* mRNA after removal of the 26-nucleotide intron results in a shift of the reading frame and continuation of the protein-coding region into the former 3' untranslated region (Fig. 3b). The protein predicted by the extended open reading frame (ORF) of the processed *XBP-1* mRNA (*XBP-1*<sup>pr</sup>) is similar in size to the 54K XBP-1 protein detected in stressed cells. To explore further the relationship between processing of the *XBP-1* mRNA and expression of the 54K protein, we modified both processed and unprocessed *XBP-1* cDNAs to encode N-terminal Flag-tagged proteins, and studied their expression in transfected cells. Human embryonic kidney (HEK) 293T cells transfected with the processed *XBP-1* cDNA constitutively expressed high levels of the 54K XBP-1



**Figure 4** Processing of *XBP-1* mRNA controls expression of the encoded protein. **a**, Immunoblot of N-terminal Flag-tagged XBP-1 from untreated or tunicamycin-treated (Tm) untransfected (UT) cells and cells transfected with expression plasmids containing the processed (*XBP-1*<sup>pr</sup>) or unprocessed (*XBP-1*<sup>up</sup>) *XBP-1* cDNA. **b**, Northern blot of mouse *XBP-1* mRNA from the transfected cells. **c**, Autoradiogram of a pulse-chase

experiment of metabolically labelled Flag-tagged XBP-1 proteins from transfected cells, and the corresponding logarithmic plot of the radioactive protein signal intensity as a function of time. **d**, Autoradiogram of metabolically labelled endogenous XBP-1 proteins from untreated, tunicamycin-treated (Tm) and proteasome-inhibited (PI) wild-type and *IRE1* mutant mouse fibroblasts.

protein (Fig. 4a). By contrast, cells transfected with *XBP-1<sup>up</sup>* expressed only low levels of a Flag-tagged protein of 33K, consistent in size with the ORF specified by the unprocessed mRNA. On treatment with tunicamycin, cells transfected with *XBP-1<sup>up</sup>* expressed high levels of the Flag-tagged 54K protein (Fig. 4a). These observations indicate that the protein encoded by the unprocessed mRNA was poorly expressed and that processing of the exogenous mRNA by endogenous IRE1 led to the expression of the 54K XBP-1 protein.

We performed pulse-chase experiments to explore the basis of the differences in expression of proteins encoded by the transfected *XBP-1<sup>up</sup>* and *XBP-1<sup>pr</sup>* mRNAs, as northern blot analysis showed similar levels of the two mRNAs (Fig. 4b). Using a short (10 min) labelling pulse, we found that the level of 54K XBP-1 accumulation was 4-fold higher than that for 33K XBP-1; the half-life of the 54K XBP-1 protein was twice that of 33K XBP-1 (22 min and 11 min, respectively). On the basis of these measurements, the synthesis rate of the 54K XBP-1 protein is 3.46-fold higher than that of 33K XBP-1. Furthermore, endogenous 54K XBP-1 accumulated to high levels in pulse-labelled wild-type cells that were treated with tunicamycin, whereas labelled 33K XBP-1 protein was barely detectable (Fig. 4d). The 33K XBP-1 signal was more obvious in IRE1 mutant cells, presumably because they accumulate *XBP-1<sup>up</sup>* mRNA (Fig. 2c). Levels of the 33K XBP-1 protein increased further in cells treated with a proteasome inhibitor (Fig. 4d, see also Supplementary Information Fig. 2), consistent with the short half-life of the protein. The half-life of endogenous 54K XBP-1 protein was approximately 21 min, similar to that of the over-expressed protein; however, the weak incorporation of label precluded accurate measurement of the half-life of endogenous 33K XBP-1 protein (Supplementary Information Fig. 2).

As in the regulation of *HAC1* in yeast, mammalian IRE1 acts as a site-specific endonuclease to cleave the mRNA of a bZIP transcription factor. Cleavage and removal of a small intron is followed by re-ligation of the 5' and 3' fragments to produce a processed mRNA that is translated more efficiently and encodes a more stable protein. These findings provide new insight into the organization and evolution of the metazoan UPR. In yeast, virtually all genes activated in the UPR are IRE1- and HAC1-dependent<sup>6,7</sup>. Nematodes lacking *ire-1* or *xbp-1* function are also unable to induce *hsp-4* (*BiP*) and other target genes of the UPR (Fig. 1j and data not shown). By contrast, IRE1 mutant mammalian cells have no obvious defect in inducing UPR markers such as *CHOP* (C/EBP homologous protein) or *XBP-1* itself (Fig. 2c). It seems that mammalian evolution was associated with specialization of IRE1 and XBP-1 function, diverting these proteins from control of many UPR target genes towards certain specific cellular functions. The phenotype of the *XBP-1<sup>-/-</sup>* mouse may provide a clue to the nature of that specialization. *RAG1<sup>-/-</sup>* (recombinase-activating gene 1) mice whose lymphoid system has been reconstituted with *XBP-1<sup>-/-</sup>* cells have a profound defect in immunoglobulin secretion, which reflects impaired differentiation of B cells to plasma cells<sup>4</sup>. The distinction between a plasma cell and B cell is based on acquisition of the ability to secrete large amounts of immunoglobulin and its morphological corollary, an elaborate ER. Our finding that IRE1 controls XBP-1 expression suggests, therefore, that increased load of client proteins in the ER activates XBP-1 and triggers development of an elaborate secretory apparatus—the hallmark of a plasma cell. □

## Methods

### Analysis of the UPR in *C. elegans*

A 1.1-kilobase (kb) fragment of *C. elegans* genomic DNA immediately 5' of the predicted initiation ATG codon of *hsp-4* was amplified by PCR and ligated in-frame with GFP in the plasmid pPD95.75 (gift of A. Fire). We generated the *hsp-4::gfp(zcls4)* V strain by co-injecting the *hsp-4::gfp* clone with the *lin-15* rescuing plasmid, pSK1 (ref. 19), into *lin-15(n765ts)* animals and then integrating the extrachromosomal array with ultraviolet/

trimethylpsoralen treatment. We screened the F<sub>2</sub> progeny of ethylmethane sulphonate (EMS)-treated *hsp-4::gfp(zcls4)* V animals for mutants with high levels of GFP expression to identify mutations that constitutively induce the UPR. Several mutations were recovered, including *upr-1(zc6)X*, which induced a high level of GFP expression in the posterior gut. To isolate mutations that blocked induction of *hsp-4::gfp*, we screened the F<sub>2</sub> progeny of EMS-treated *hsp-4::gfp(zcls4)* V; *upr-1(zc6)X* animals for mutants with a reduced GFP signal. We mapped the mutations to visible markers and single nucleotide polymorphic markers.

RNAi was performed by injecting double-stranded RNA—made *in vitro* through using as a template the 720-bp *SalI*–*EcoRI* fragment of the *ire-1* cDNA (clone yk8e9) or the 515-bp *SalI*–*NcoI* fragment of the *xbp-1* cDNA (clone yk146d1)—into young adult hermaphrodites, and then observing the phenotype of the F<sub>1</sub> progeny as described previously<sup>20</sup>. We transferred animals to plates containing 5 µg ml<sup>-1</sup> tunicamycin (Calbiochem), and visualized the GFP expression using an epifluorescent stereomicroscope (Zeiss M2 Bio). RNA for northern blot analysis was isolated by acid-guanidinium thiocyanate-phenol-chloroform extraction. We used a radiolabelled 100-bp *HindIII*–*XhoI* fragment from *hsp-4* cDNA clone yk34e10 as a probe.

### Cell culture, transfection, immunoblot and immunoprecipitation

Mouse embryonic fibroblasts and embryonic stem cells lacking IRE1 activity (*IRE1α<sup>-/-</sup>* or *IRE1α; IRE1β<sup>-/-</sup>*), and *PERK<sup>-/-</sup>* fibroblasts were cultured as described previously<sup>9,12</sup>. Cells were treated with tunicamycin (2.5 µg ml<sup>-1</sup>; Calbiochem), thapsigargin (100 nM; Sigma), dithiothreitol (2 mM; Sigma) or the proteasome inhibitor MG132 (25 µM; Calbiochem), as indicated. To activate B-cell differentiation *in vitro* cells were teased from crushed adult mouse spleen and cultured in the presence of 25 µg ml<sup>-1</sup> *Escherichia coli* lipopolysaccharide.

We purchased a full-length murine *XBP-1* cDNA (expressed sequence tag BF454459, Research Genetics). The processed version of the cDNA was created by replacing the *XhoI*–*EcoRV* fragment of the unprocessed cDNA with the equivalent fragment from the processed cDNA obtained by RT-PCR of mRNA from stressed cells. Flag-tagged versions of the two cDNAs were constructed by inserting the Flag epitope tag sequence immediately 3' of the ATG start codon, and the cDNAs were expressed from a plasmid containing the human cytomegalovirus (CMV) promoter (pFlag-CMV2).

XBP-1 proteins in whole-cell extracts were detected by an immunoblot of 9% SDS–polyacrylamide gel electrophoresis (PAGE) using a rabbit polyclonal serum directed to residues 97–267, common to both 54K and 33K XBP-1 proteins (sc-8015; Santa Cruz Biotechnology). XBP-1 proteins in transfected cells were detected by a monoclonal antibody to the Flag epitope tag (Kodak-IBI). The transcription factors CREB and CHOP were detected as described previously<sup>9</sup>.

Newly synthesized proteins were labelled *in vivo* with <sup>35</sup>S-methionine and <sup>35</sup>S-cysteine (500 µCi ml<sup>-1</sup>, Translabel, ICN Biochemicals). After removal of the labelling medium cells were washed and incubated in complete medium for cold chase. Labelled XBP-1 proteins from whole-cell extracts prepared in radio-immunoprecipitation buffer<sup>9</sup> were immunoprecipitated using the anti-Flag or anti-XBP-1 antibodies described above and revealed by autoradiography after 9% SDS–PAGE. Incorporation of the radiolabel was quantified by phosphorimager analysis. In calculating the relative synthesis rates of 54K XBP-1 and 33K XBP-1, the relative incorporation of label was corrected for differences in the content of methionine (7:6) and cysteine (4:3) predicted for the two proteins.

### Analysis of XBP-1 mRNA cleavage by IRE1

RNA from untreated and tunicamycin-treated wild-type and *IRE1α<sup>-/-</sup>* mouse embryonic stem cells was reverse transcribed using an *XBP-1*-specific antisense primer, mXBP1.4AS: 5'-GCACCTTCTAGAAAGCTACACTAGCA-3'. Nested PCR using the sense primer mXBP1.3S (5'-AAACAGAGTAGCAGCGCAGACTGC-3') and the antisense primer mXBP1.2AS: (5'-GGATCTCTAAAAGTCTAGAGGCTTGTTG-3') amplified a 600-bp cDNA product encompassing the IRE1 cleavage sites. This fragment was further digested by *PstI* to reveal a restriction site that is lost after IRE1-mediated cleavage and splicing of the mRNA. The cDNA fragments were resolved on a 2% agarose gel and revealed by Southern blot hybridized to the <sup>32</sup>P-labelled *XhoI*–*PstI* fragment of the unprocessed *XBP-1* cDNA. The processed cDNAs from mouse and *C. elegans* were sequenced and the sequence deposited in GenBank under accession numbers AF443192 and AF443191, respectively.

The *XhoI*–*EcoRV* fragment of the *XBP-1* cDNA from unstressed (*XBP-1<sup>up</sup>*) and stressed (*XBP-1<sup>pr</sup>*) mouse embryonic stem cells was ligated into pBluescript plasmid (Stratagene) and sense strand <sup>32</sup>P-labelled RNA was transcribed *in vitro* using a kit from Promega. The labelled RNA was incubated in the presence or absence of 1 mM ATP with the recombinant cytoplasmic domain of IRE1β purified from SF9 cells as described<sup>21</sup>. The radiolabelled RNA fragments were resolved on 8 M urea 6% PAGE gel and revealed by autoradiography.

Received 28 August; accepted 23 November 2001.

1. Mori, K. Tripartite management of unfolded proteins in the endoplasmic reticulum. *Cell* **101**, 451–454 (2000).
2. Kaufman, R. J. Stress signaling from the lumen of the endoplasmic reticulum: coordination of gene transcriptional and translational controls. *Genes Dev.* **13**, 1211–1233 (1999).
3. Shamu, C. E. Splicing: HACKing into the unfolded-protein response. *Curr. Biol.* **8**, R121–R123 (1998).
4. Reimold, A. M. *et al.* Plasma cell differentiation requires the transcription factor XBP-1. *Nature* **412**, 300–307 (2001).
5. Claus, I. M. *et al.* *In situ* hybridization studies suggest a role for the basic region-leucine zipper protein hXBP-1 in exocrine gland and skeletal development during mouse embryogenesis. *Dev. Dyn.* **197**, 146–156 (1993).
6. Travers, K. J. *et al.* Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation. *Cell* **101**, 249–258 (2000).

7. Casagrande, R. *et al.* Degradation of proteins from the ER of *S. cerevisiae* requires an intact unfolded protein response pathway. *Mol. Cell* **5**, 729–735 (2000).
8. Scheuner, D. *et al.* Translational control is required for the unfolded protein response and *in vivo* glucose homeostasis. *Mol. Cell* **7**, 1165–1176 (2001).
9. Harding, H. *et al.* Regulated translation initiation controls stress-induced gene expression in mammalian cells. *Mol. Cell* **6**, 1099–1108 (2000).
10. Yoshida, H. *et al.* ATF6 activated by proteolysis binds in the presence of NF-Y (CBF) directly to the *cis*-acting element responsible for the mammalian unfolded protein response. *Mol. Cell Biol.* **20**, 6755–6767 (2000).
11. Ye, J. *et al.* ER stress induces cleavage of membrane-bound ATF6 by the same proteases that process SREBPs. *Mol. Cell* **6**, 1355–1364 (2000).
12. Urano, F. *et al.* Coupling of stress in the endoplasmic reticulum to activation of JNK protein kinases by transmembrane protein kinase IRE1. *Science* **287**, 664–666 (2000).
13. Yoshida, H., Haze, K., Yanagi, H., Yura, T. & Mori, K. Identification of the *cis*-acting endoplasmic reticulum stress response element responsible for transcriptional induction of mammalian glucose-regulated proteins. Involvement of basic leucine zipper transcription factors. *J. Biol. Chem.* **273**, 33741–33749 (1998).
14. Tirasophon, W., Welihinda, A. A. & Kaufman, R. J. A stress response pathway from the endoplasmic reticulum to the nucleus requires a novel bifunctional protein kinase/endoribonuclease (Ire1p) in mammalian cells. *Genes Dev.* **12**, 1812–1824 (1998).
15. Bertolotti, A. *et al.* Increased sensitivity to dextran sodium sulfate colitis in IRE1b deficient mice. *J. Clin. Invest.* **107**, 585–593 (2001).
16. Wang, X. Z. *et al.* Cloning of mammalian Ire1 reveals diversity in the ER stress responses. *EMBO J.* **17**, 5708–5717 (1998).
17. Gonzalez, T. N., Sidrauski, C., Dorfner, S. & Walter, P. Mechanism of non-spliceosomal mRNA splicing in the unfolded protein response pathway. *EMBO J.* **18**, 3119–3132 (1999).
18. Kawahara, T., Yanagi, H., Yura, T. & Mori, K. Unconventional splicing of HAC1/ERN4 mRNA required for the unfolded protein response. Sequence-specific and non-sequential cleavage of the splice sites. *J. Biol. Chem.* **273**, 1802–1807 (1998).
19. Clark, S. G., Lu, X. & Horvitz, H. R. The *Caenorhabditis elegans* locus lin-15, a negative regulator of a tyrosine kinase signaling pathway, encodes two different proteins. *Genetics* **137**, 987–997 (1994).
20. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
21. Niwa, M., Sidrauski, C., Kaufman, R. J. & Walter, P. A role for presenilin-1 in nuclear accumulation of Ire1 fragments and induction of the mammalian unfolded protein response. *Cell* **99**, 691–702 (1999).

**Supplementary Information** accompanies the paper on Nature's website (<http://www.nature.com>).

## Acknowledgements

We thank C. Chiu for help in initiating the nematode work; A. Bertolotti, D. Levy and E. Y. Skolnik for useful discussions; Y. Kohara for *C. elegans* EST clones; and A. Ron and R. Onn for calculating the synthesis rate of XBP1 proteins. This work was supported by grants from the National Institutes of Health. D.R. is a Scholar of the Ellison Medical Foundation.

## Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.R. (e-mail: [ron@saturn.med.nyu.edu](mailto:ron@saturn.med.nyu.edu)).

# Superoxide activates mitochondrial uncoupling proteins

**Karim S. Echtay\***, **Damien Roussel\***, **Julie St-Pierre\***, **Mika B. Jakabsons\***, **Susana Cadenas\***, **Jeff A. Stuart\***, **James A. Harper\***, **Stephen J. Roebuck\***, **Alastair Morrison†**, **Susan Pickering‡**, **John C. Clapham‡** & **Martin D. Brand\***

\* Medical Research Council Dunn Human Nutrition Unit, Hills Road, Cambridge CB2 2XY, UK  
Departments of †Comparative Genomics, and ‡ Vascular Biology, GlaxoSmithKline, Harlow, Essex CM19 5AW, UK

**Uncoupling protein 1 (UCP1) diverts energy from ATP synthesis to thermogenesis in the mitochondria of brown adipose tissue by catalysing a regulated leak of protons across the inner membrane<sup>1,2</sup>. The functions of its homologues, UCP2 and UCP3, in other tissues are debated<sup>3,4</sup>. UCP2 and UCP3 are present at much lower abundance than UCP1, and the uncoupling with**

**which they are associated is not significantly thermogenic<sup>5,6</sup>. Mild uncoupling would, however, decrease the mitochondrial production of reactive oxygen species, which are important mediators of oxidative damage<sup>7,8</sup>. Here we show that superoxide increases mitochondrial proton conductance through effects on UCP1, UCP2 and UCP3. Superoxide-induced uncoupling requires fatty acids and is inhibited by purine nucleotides. It correlates with the tissue expression of UCPs, appears in mitochondria from yeast expressing UCP1, and is absent in skeletal muscle mitochondria from UCP3 knockout mice. Our findings indicate that the interaction of superoxide with UCPs may be a mechanism for decreasing the concentrations of reactive oxygen species inside mitochondria.**

As coenzyme Q (CoQ) has been identified as a regulatory cofactor for proton transport by UCP1 (ref. 9), UCP2 and UCP3 (ref. 10) in liposomes, we tested the effect of CoQ in isolated mitochondria. CoQ increased proton conductance in rat kidney (but not liver) mitochondria that were oxidizing succinate. This increase required fatty acids and was prevented by GDP. CoQ activated proton conductance only when it was likely to be reduced to CoQH<sub>2</sub>. Activation was abolished by superoxide dismutase, indicating that CoQ might mediate uncoupling through the production of superoxide<sup>11</sup>. To explore this possibility, we replaced CoQ by xanthine plus xanthine oxidase—an exogenous system that generates superoxide. Proton conductance increased, indicating that CoQ acted in mitochondria through the production of superoxide.

Incubating rat skeletal muscle mitochondria with xanthine plus xanthine oxidase to generate superoxide increased proton conductance (Fig. 1a). This is seen as an increased rate of proton leak at each membrane potential, resulting in a curve that is deflected upwards. This increase was fully inhibited either by superoxide dismutase, indicating that it was dependent on superoxide, or by 500 μM GDP (Fig. 1a). GDP had no effect on control mitochondria (data not shown), confirming previous results<sup>12</sup>.

The superoxide effect was abolished by bovine serum albumin (BSA), which binds fatty acids (Fig. 1b), but restored by adding palmitic acid in the presence of BSA (Fig. 1c), indicating that activation by superoxide requires fatty acids. Proton conductance that is activated by fatty acids and sensitive to GDP is characteristic of uncoupling by UCP1 in brown adipose tissue (BAT) mitochondria<sup>12</sup>, suggesting that the uncoupling caused by superoxide in skeletal muscle mitochondria (which lack UCP1 and -2 (ref. 13) but contain UCP3) was mediated by UCP3.

We also investigated skeletal muscle mitochondria from starved rats. Starvation for 24 h doubles the concentration of UCP3 protein without affecting the basal level of proton conductance<sup>12</sup>. Superoxide stimulated proton conductance twice as strongly in mitochondria from starved rats (Fig. 1d) as in mitochondria from fed rats (Fig. 1a). The same was true in the presence of BSA plus palmitate (data not shown). This correlates with a near doubling of UCP3 protein in starved rats, which was confirmed by western blot (data not shown), and implicates UCP3 in the superoxide effect.

Confirmation of the role of UCP3 was obtained using skeletal muscle mitochondria isolated from UCP3 knockout mice, which had the same basal proton conductance as the controls. Muscle mitochondria from wild-type mice showed the same GDP-sensitive stimulation of proton conductance as those from rats (Fig. 1e) and the same dependence on fatty acids (data not shown). However, superoxide had no effect on mitochondria from the skeletal muscle of UCP3 knockout mice (Fig. 1f), showing that superoxide uncoupled wild-type mitochondria by interacting with UCP3. We verified that the lack of effect of xanthine plus xanthine oxidase in UCP3 knockouts was not caused by lack of superoxide by a direct assay of superoxide production using superoxide dismutase and a homovanillic acid/horseradish peroxidase fluorescence assay<sup>14</sup> (data not shown). Thus, the fatty-acid-dependent, GDP-sensitive increase in proton conductance caused by xanthine plus xanthine oxidase



was caused by an effect of superoxide acting through UCP3.

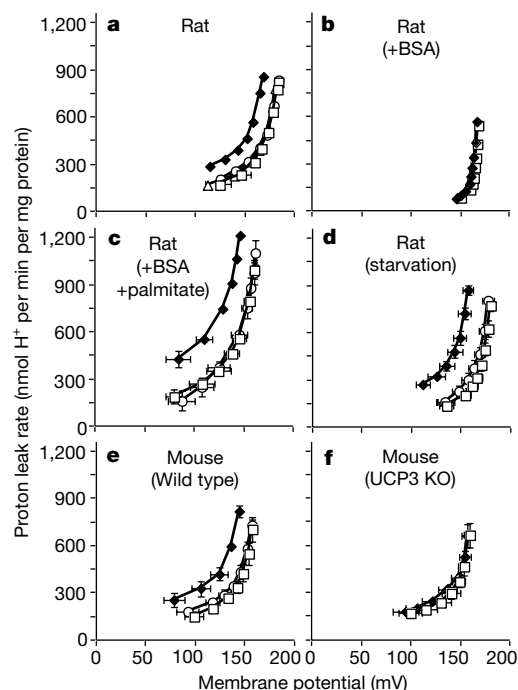
We investigated whether UCP2 was also activated by superoxide. Because UCP2 is expressed widely<sup>4,13</sup>, we checked the superoxide effect on basal proton conductance in mitochondria from several tissues that do not express UCP3 (ref. 4). Superoxide increased proton conductance in mitochondria from rat kidney in a GDP-sensitive manner (Fig. 2a). This activation required fatty acids (Fig. 2b, c). But in liver and heart mitochondria, which lack UCP2 (ref. 13), basal proton leak was not stimulated by superoxide (Fig. 2d, e). Mitochondria from spleen<sup>13</sup> and pancreatic  $\beta$  cells<sup>15</sup> do contain UCP2, and superoxide increased the proton conductance in the absence, but not the presence, of GDP in these mitochondria (Fig. 2f, g). These results indicate that superoxide probably acts through UCP2 in mitochondria from the kidney, spleen and pancreatic  $\beta$  cells.

To determine whether UCP1 is also activated by superoxide, we examined BAT mitochondria from warm-adapted rats. These mitochondria contain relatively low amounts of UCP1, which allows coupling to be measured in the absence of GDP. BAT mitochondria had low endogenous proton conductance through UCP1, which was reduced to the basal level by adding GDP (Fig. 3a). The endogenous UCP1 activity was stimulated by superoxide, and this stimulation was prevented by superoxide dismutase. GDP abolished both the endogenous activity and the superoxide-stimulated activity, returning proton conductance to the basal level.

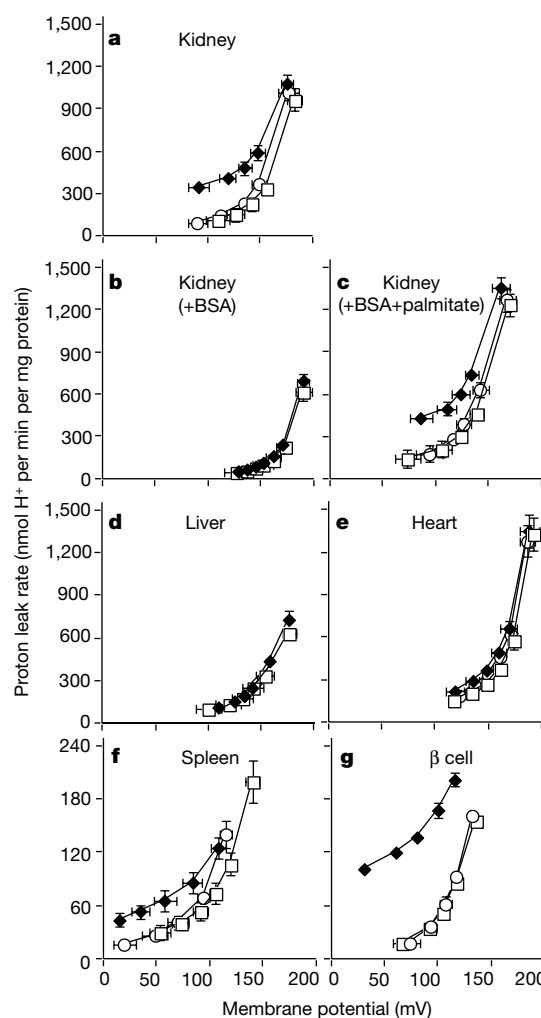
Thus, BAT mitochondria showed the same response to superoxide as skeletal muscle and kidney mitochondria, but it was

superimposed on the normal GDP-sensitive uncoupling mediated by UCP1. Unlike the effects in other tissues, superoxide was able to uncouple BAT mitochondria even in the presence of BSA, indicating either that UCP1 has a higher affinity for fatty acids than UCP2 and -3, or that BAT mitochondria have more contaminating free fatty acids than other tissues. These results suggest that UCP1 also facilitates superoxide uncoupling; however, BAT mitochondria may also contain both UCP2 and -3 (ref. 4), which complicates the interpretation.

To establish whether UCP1 was responsible for the superoxide effects in BAT mitochondria, we expressed mouse UCP1 in the yeast *Saccharomyces cerevisiae*, which has been previously shown to provide a convenient model for studying UCP1 function<sup>16–18</sup>. Mitochondria from control yeast hardly responded to superoxide or GDP (Fig. 3b); in contrast, mitochondria from yeast expressing modest concentrations of UCP1 ( $\sim 1 \mu\text{g}$  per mg protein)<sup>18</sup> had greater proton conductance than control yeast mitochondria, and this extra proton conductance was completely sensitive to GDP



**Figure 1** Effect of superoxide on the proton conductance of skeletal muscle mitochondria: superoxide activation of UCP3. Graphs show rate of proton leak as a function of its driving force (the mitochondrial membrane potential) as the potential was varied by titration with succinate. Open squares, control; filled diamonds, 50  $\mu\text{M}$  xanthine plus xanthine oxidase (0.01 U per 3.5 ml) added before TPMP<sup>+</sup>; open circles, xanthine plus xanthine oxidase and 500  $\mu\text{M}$  GDP added before TPMP<sup>+</sup>; open triangles, xanthine plus xanthine oxidase and 12 U ml<sup>-1</sup> superoxide dismutase added before TPMP<sup>+</sup>. Mitochondria were from control rats fed *ad libitum* (a); control rats, in medium supplemented with 0.3% defatted BSA (b); control rats, in medium with 0.3% BSA and 300  $\mu\text{M}$  palmitic acid (c); rats starved for 24 h (d); wild-type mice (e); and littermate homozygous UCP3 knockout mice (f). Data are the mean  $\pm$  s.e.m. of three independent experiments each performed in duplicate.



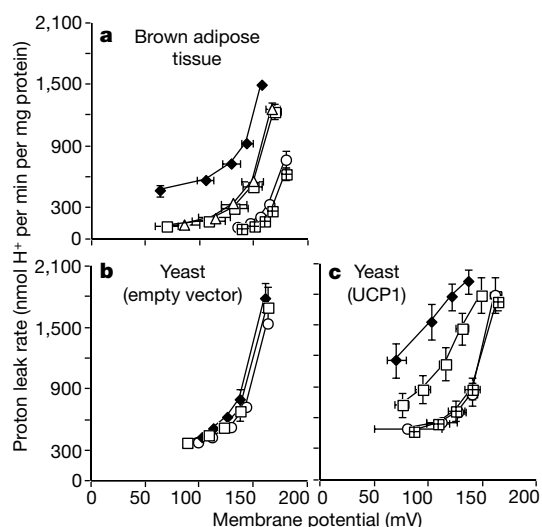
**Figure 2** Effect of superoxide on the proton conductance of mitochondria from different tissues: superoxide activation of UCP2. The potential was varied by adding cyanide (up to 100  $\mu\text{M}$ ) using 4 mM succinate as the substrate. Open squares, control; filled diamonds, 50  $\mu\text{M}$  xanthine plus xanthine oxidase (0.01 U per 3.5 ml) added before TPMP<sup>+</sup>; open circles, xanthine plus xanthine oxidase and 500  $\mu\text{M}$  GDP added before TPMP<sup>+</sup>. Mitochondria were from rat kidney (a); rat kidney, in medium with 0.3% BSA (b); rat kidney, in medium with 0.3% BSA and 300  $\mu\text{M}$  palmitic acid (c); rat liver (d); rat heart (e); rat spleen (f); mouse min6 pancreatic  $\beta$  cells (g). In a–f, data are the mean  $\pm$  s.e.m. of three independent experiments each performed in duplicate; in g, data are the mean  $\pm$  range of duplicate experiments using a single preparation.

(Fig. 3c). Thus, mouse UCP1 showed native uncoupling behaviour in yeast at these levels of expression. Superoxide stimulated the proton conductance and GDP returned it to the basal level (Fig. 3c). These results clearly show that superoxide increases GDP-sensitive proton conductance through UCP1.

UCP1 accepts many nucleotides but has a strong preference for purine nucleoside diphosphates and triphosphates<sup>2</sup>. We determined the nucleotide specificity and binding affinity of the inducible proton conductance catalysed by UCP2 (kidney mitochondria) and -3 (skeletal muscle mitochondria). These proteins had similar nucleotide specificity, with potent inhibition by purine but not pyrimidine nucleoside diphosphates and triphosphates (Fig. 4a). One difference was that CTP, CDP and perhaps UTP partially inhibited UCP2 but not -3. Nucleoside monophosphates had no effect on either protein. AMP had a separate stimulatory effect on skeletal muscle (but not kidney) mitochondrial proton conductance, which is mediated through the adenine nucleotide translocase, as reported previously<sup>19</sup>. The inhibition by GDP was found to follow simple saturation kinetics (Fig. 4b). Superoxide activation was inhibited by 50% with  $17 \pm 0.9 \mu\text{M}$  GDP in kidney and  $8.5 \pm 0.5 \mu\text{M}$  in skeletal muscle mitochondria.

Stimulation of proton conductance by xanthine plus xanthine oxidase was insensitive to catalase (Fig. 4a), showing that it was not caused by hydrogen peroxide. Kidney or liver mitochondria incubated with glucose plus glucose oxidase to generate exogenous  $\text{H}_2\text{O}_2$  had increased basal proton conductance that was insensitive to GDP (data not shown), suggesting that peroxide or its by-products damaged the membrane and increased its proton permeability but did not activate UCPs.

Stimulation by superoxide was not inhibited by glybenclamide, a  $\text{K}_{\text{ATP}}$  channel blocker; by bongkrekate or cyclosporin A, inhibitors of the mitochondrial permeability transition; or by carboxyatractylate or bongkrekate, inhibitors of the adenine nucleotide translocase (Fig. 4a). Activation by superoxide required a few minutes to become maximal (possibly because superoxide must react or be translocated before uncoupling). The superoxide effect was fully



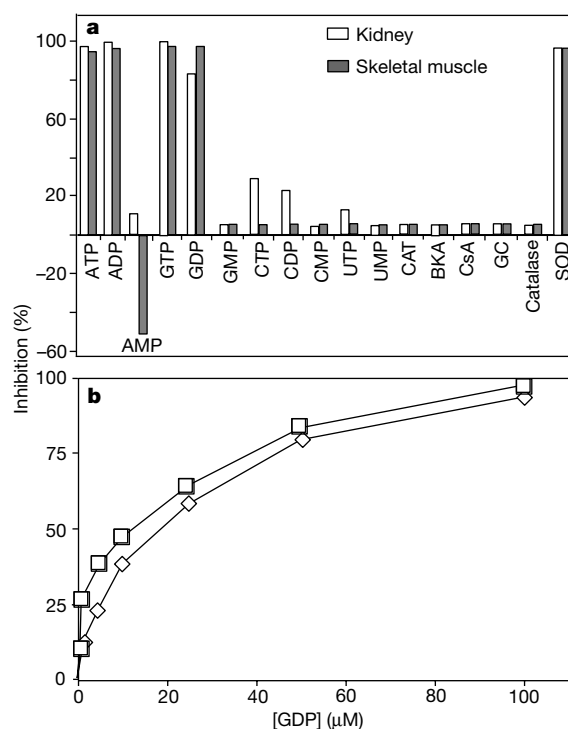
**Figure 3** Effect of superoxide on the proton conductance of mitochondria from brown adipose tissue and transgenic yeast: superoxide activation of UCP1. Open squares, control; crossed squares, control plus 500  $\mu\text{M}$  GDP added before  $\text{TPMP}^+$ ; filled diamonds, 50  $\mu\text{M}$  xanthine plus xanthine oxidase (0.01 U per 3.5 ml) added before  $\text{TPMP}^+$ ; open circles, xanthine plus xanthine oxidase and 500  $\mu\text{M}$  GDP added before  $\text{TPMP}^+$ ; open triangles, xanthine plus xanthine oxidase and 12 U  $\text{ml}^{-1}$  superoxide dismutase added before  $\text{TPMP}^+$ . Proton leak in mitochondria from brown adipose tissue (a); yeast containing control empty vector (b); yeast expressing mouse UCP1 from pBF307 vector<sup>18</sup> (c). Data are the mean  $\pm$  s.e.m. (or range) of 2–3 independent experiments each performed in duplicate.

reversible, because the proton conductance was the same as controls after a 10-min incubation of mitochondria with xanthine plus xanthine oxidase followed by the addition of superoxide dismutase.

Activation in kidney mitochondria was less at pH 6.5 and greater at pH 7.8 than at pH 7.2 (data not shown), perhaps because of increased protonation and dismutation (and hence destruction) of superoxide at more acid pH (ref. 20). GDP sensitivity was abolished at pH 7.8, however, indicating that nucleotide inhibition of UCP2 may be very sensitive to pH, as it is in UCP1 (ref. 2). Oxygen consumption by kidney mitochondria incubated with xanthine, xanthine oxidase and myxothiazol was very low, showing that superoxide does not generate its effects by causing electron flow to bypass complex III of the respiratory chain.

We conclude that superoxide interacts with UCP1, -2 and -3, which leads to an increase in proton conductance that requires fatty acids and is inhibited by purine nucleotides. This extends previous findings that reactive oxygen species (ROS) also cause fatty-acid-dependent, ATP-inhibited uncoupling in plant mitochondria<sup>21,22</sup>. Because our studies were performed in isolated mitochondria in the presence of high concentrations of exogenous superoxide, we do not know whether superoxide-stimulated UCP-mediated uncoupling occurs in cells or in intact organisms. This will need to be addressed by future experiments.

Two types of model could explain our results: the effects of superoxide might be direct, or might take place through some product (other than peroxide). In type I models, superoxide induces proton transport by allosterically activating the proton transport mechanisms previously proposed for UCP1 (refs 2, 23). In type II models, UCPs use the mitochondrial membrane potential to export endogenously produced superoxide anions from the matrix to the



**Figure 4** Nucleotide specificity and affinity of UCP2 (kidney) and UCP3 (skeletal muscle). Inhibition of proton conductance in standard medium is presented as the percentage inhibition of superoxide-activated  $\text{H}^+$  transport rates at defined membrane potential (130 mV, interpolated from proton leak curves). a, Nucleotides were added at 100  $\mu\text{M}$ , carboxyatractylate (CAT) at 1.5  $\mu\text{M}$ , bongkrekate (BKA) at 1.5  $\mu\text{M}$ , cyclosporin A (CsA) at 1  $\mu\text{M}$ , glybenclamide (GC) at 1  $\mu\text{M}$ , catalase at 3 U  $\text{ml}^{-1}$  and superoxide dismutase (SOD) at 12 U  $\text{ml}^{-1}$ . b, Dependence of inhibition on concentration of GDP. Open squares, skeletal muscle; open diamonds, kidney. Data are the mean and range of two determinations.

intermembrane space where they can either undergo more rapid dismutation because of the more acid pH or be scavenged by cytochrome *c*, CoQ or other antioxidant defences. Uncoupling by exogenous superoxide would be the result of protonation in the intermembrane space (the pK is 4.8)<sup>20</sup>, followed by hydroperoxyl-radical diffusion into the matrix and export of superoxide anions by UCP—a similar pathway to one proposed previously<sup>24</sup>. In this model, UCPs would normally export superoxide produced in the matrix and would only show up as an uncoupling pathway when the matrix was flooded with superoxide being produced exogenously and continuously (for example, by xanthine oxidase).

In both types of model, UCPs can also decrease ROS production by mild uncoupling, perhaps as a feedback response to the overproduction of ROS by the electron transport chain. The models can explain the increased production of matrix ROS observed in mitochondria and tissues from UCP knockout mice<sup>5,6</sup>. The function of UCP2 and -3 might therefore be more associated with protection against ROS than with thermogenesis. This would explain the induction of UCP expression by cold in plants<sup>25</sup>, which has been postulated to increase the production of ROS<sup>26</sup>.

We speculate that UCP1 evolved a thermogenic role in mammals as a side pathway of an original, more general function of protection against the cold-induced production of ROS. Such a function can explain the occurrence of UCP2 and -3 in ectotherms<sup>3</sup>, the association of UCP2 with cells of the immune system in mammals<sup>5,13</sup>, and the observation of nucleotide-sensitive ROS production in cells expressing UCP2 (ref. 27). It also might explain why UCP2 knockout mice are resistant to infection by endoparasites<sup>5</sup>, as they would lack a system that normally removes ROS, and suggests interpretations of the observation that these animals have altered pancreatic  $\beta$ -cell function<sup>15</sup>. □

## Methods

### Proton leak measurements

We measured respiration rate and membrane potential simultaneously by using electrodes sensitive to oxygen and to the potential-dependent probe triphenylmethyl phosphonium cation (TPMP<sup>+</sup>)<sup>28</sup>. The TPMP<sup>+</sup>-binding correction was assumed to be 0.4/( $\mu$ l per mg protein). Proton leak rates were calculated by multiplying oxygen consumption rates by the H<sup>+</sup>/O ratio of six.

Mitochondria were prepared essentially as described<sup>29</sup>. Mitochondria isolated from total hindlimb skeletal muscle (0.35 mg protein ml<sup>-1</sup>), liver (0.5 mg ml<sup>-1</sup>), kidney (0.35 mg ml<sup>-1</sup>), heart (0.35 mg ml<sup>-1</sup>), spleen (1.0 mg ml<sup>-1</sup>) or  $\beta$  cells (0.75 mg ml<sup>-1</sup>) were incubated in standard assay medium containing 120 mM KCl, 5 mM KH<sub>2</sub>PO<sub>4</sub>, 3 mM HEPES and 1 mM EGTA (pH 7.2 and 37 °C) with 5  $\mu$ M rotenone (a complex I inhibitor to prevent oxidation of any endogenous NAD-linked substrates), 80 ng nigericin ml<sup>-1</sup> (to abolish the pH gradient) and 1  $\mu$ g oligomycin ml<sup>-1</sup> (to prevent ATP synthesis). We calibrated the TPMP<sup>+</sup> electrode with sequential 1- $\mu$ M additions of TPMP<sup>+</sup> to 5  $\mu$ M. Skeletal muscle mitochondria were titrated by incremental additions of succinate to 1 mM (oxidizable substrate); other tissues were titrated by cyanide (up to 100  $\mu$ M) using 4 mM succinate as the substrate. After each run, 0.2  $\mu$ M FCCP (carbonylcyanide *p*-trifluoromethoxyphenylhydrazone) was added to release TPMP<sup>+</sup> for baseline correction. The  $\beta$  cell line was kindly provided by F. Gribble and grown as described<sup>30</sup>.

Proton leak in BAT mitochondria (0.35 mg protein ml<sup>-1</sup>) isolated from rats maintained at 25 °C was measured in assay medium containing 50 mM KCl, 5 mM HEPES, 1 mM EGTA, 4 mM KH<sub>2</sub>PO<sub>4</sub> (pH 7.2 and 37 °C) and supplemented with 1% BSA. BAT mitochondria were titrated by cyanide (up to about 100  $\mu$ M) using 10 mM  $\alpha$ -glycerophosphate as the substrate.

Proton leak in yeast mitochondria<sup>18</sup> (0.4 mg protein ml<sup>-1</sup>) was measured in assay medium containing 10 mM Tris-maleate, 650 mM sorbitol, 0.5 mM EGTA, 2 mM MgCl<sub>2</sub>, 10 mM K<sub>2</sub>HPO<sub>4</sub> (pH 6.8 and 30 °C). The TPMP<sup>+</sup> electrode was calibrated with sequential 1- $\mu$ M additions of TPMP<sup>+</sup> to 4  $\mu$ M, and then oligomycin (10  $\mu$ g ml<sup>-1</sup>) and nigericin (100 ng ml<sup>-1</sup>) in ethanol were added. Oxidation of this ethanol (0.2%) was inhibited progressively through successive steady states by adding cyanide up to about 50  $\mu$ M.

### Generation of UCP3 knockout mice

We constructed a targeting deletion vector containing a neomycin cassette flanked by a 3.2-kilobase (kb) *NheI*–*SacI* fragment of the mouse *ucp3* gene for 3' homology and a 3-kb *NheI*–*EcoRI* fragment for 5' homology. The 8.7-kb targeting construct was designed to

generate a 2.5-kb deletion to remove exons 1 and 2 (exon 2 contains the start codon) from the mouse *ucp3* gene. Neomycin-resistant clones were isolated and injected into C57BL/6 $\times$  CBA F<sub>2</sub> embryos. These chimeric mice were backcrossed for six generations onto a C57BL/6 background (N6). Western blot analysis of mitochondria and northern blot analysis of RNA from skeletal muscle of wild-type and *ucp3*<sup>-/-</sup> mice confirmed that there was no UCP3 expression in the knockouts.

Received 19 July; accepted 10 October 2001.

- Nicholls, D. G. & Rial, E. A history of the first uncoupling protein, UCP1. *J. Bioenerg. Biomembr.* **31**, 399–406 (1999).
- Klingenberg, M. & Echay, K. S. Uncoupling proteins: the issues from a biochemist point of view. *Biochim. Biophys. Acta* **1504**, 128–143 (2001).
- Stuart, J. A., Cadenas, S., Jekabsons, M. B., Roussel, D. & Brand, M. D. Mitochondrial proton leak and the uncoupling protein 1 homologues. *Biochim. Biophys. Acta* **1504**, 144–158 (2001).
- Ricquier, D. & Bouillaud, F. The uncoupling protein homologues: UCP1, UCP2, UCP3, StUCP and AtUCP. *Biochem. J.* **345**, 161–179 (2000).
- Arsenijevic, D. *et al.* Disruption of the uncoupling protein-2 gene in mice reveals a role in immunity and reactive oxygen species production. *Nature Genet.* **26**, 435–439 (2000).
- Vidal-Puig, A. J. *et al.* Energy metabolism in uncoupling protein 3 gene knockout mice. *J. Biol. Chem.* **275**, 16258–16266 (2000).
- Papa, S. & Skulachev, V. P. Reactive oxygen species, mitochondria, apoptosis and aging. *Mol. Cell Biochem.* **174**, 305–319 (1997).
- Brand, M. D. Uncoupling to survive? The role of mitochondrial inefficiency in ageing. *Exp. Gerontol.* **35**, 811–820 (2000).
- Echay, K. S., Winkler, E. & Klingenberg, M. Coenzyme Q is an obligatory cofactor for uncoupling protein function. *Nature* **408**, 609–613 (2000).
- Echay, K. S., Winkler, E., Frischmuth, K. & Klingenberg, M. Uncoupling proteins 2 and 3 are highly active H<sup>+</sup> transporters and highly nucleotide sensitive when activated by coenzyme Q (ubiquinone). *Proc. Natl Acad. Sci. USA* **98**, 1416–1421 (2001).
- Echay, K. S. & Brand, M. D. Coenzyme Q induces GDP sensitive proton conductance in kidney mitochondria. *Biochem. Soc. Trans.* **29**, 763–768 (2001).
- Cadenas, S. *et al.* UCP2 and UCP3 rise in starved rat skeletal muscle but mitochondrial proton conductance is unchanged. *FEBS Lett.* **462**, 257–260 (1999).
- Pecqueur, C. *et al.* Uncoupling Protein 2, *in vivo* distribution, induction upon oxidative stress, and evidence for translational regulation. *J. Biol. Chem.* **276**, 8705–8712 (2001).
- Ruch, W., Cooper, P. H. & Baggiolini, M. Assay of H<sub>2</sub>O<sub>2</sub> production by macrophages and neutrophils with homovanillic acid and horse-radish peroxidase. *J. Immunol. Methods* **63**, 347–357 (1983).
- Zhang, C. Y. *et al.* Uncoupling protein-2 negatively regulates insulin secretion and is a major link between obesity,  $\beta$ -cell dysfunction and type 2 diabetes. *Cell* **105**, 745–755 (2001).
- Bathgate, B., Freebairn, E. M., Greenland, A. J. & Reid, G. A. Functional expression of the rat brown adipose tissue uncoupling protein in *Saccharomyces cerevisiae*. *Mol. Microbiol.* **6**, 363–370 (1992).
- Echay, K. S., Bienengraeber, M. & Klingenberg, M. Mutagenesis of the uncoupling protein of brown adipose tissue. Neutralization of E190 largely abolishes pH control of nucleotide binding. *Biochemistry* **36**, 8253–8260 (1997).
- Stuart, J. A., Harper, J. A., Jekabsons, M. B., Brindle, K. M. & Brand, M. D. A mitochondrial uncoupling artefact can be caused by expression of uncoupling protein 1 in yeast. *Biochem. J.* **356**, 779–789 (2001).
- Cadenas, S. *et al.* AMP decreases the efficiency of skeletal-muscle mitochondria. *Biochem. J.* **351**, 307–311 (2000).
- Halliwel, B. & Gutteridge, J. M. C. *Free Radicals in Biology and Medicine* 2nd edn (Clarendon, Oxford, 1989).
- Pastore, D., Fratianni, A., Di Pede, S. & Passarella, S. Effects of fatty acids, nucleotides and reactive oxygen species on durum wheat mitochondria. *FEBS Lett.* **470**, 88–92 (2000).
- Kowaltowski, A. J., Costa, A. D. & Vercesi, A. E. Activation of the potato plant uncoupling mitochondrial protein inhibits reactive oxygen species generation by the respiratory chain. *FEBS Lett.* **425**, 213–216 (1998).
- Jezek, P., Orosz, D. E., Modriansky, M. & Garlid, K. D. Transport of anions and protons by the mitochondrial uncoupling protein and its regulation by nucleotides and fatty acids. A new look at old hypotheses. *J. Biol. Chem.* **269**, 26184–26190 (1994).
- Liu, S. S. Generating, partitioning, targeting and functioning of superoxide in mitochondria. *Biosci. Rep.* **17**, 259–272 (1997).
- Laloi, M. *et al.* A plant cold-induced uncoupling protein. *Nature* **389**, 135–136 (1997).
- Scandalios, J. G. Oxygen stress and superoxide dismutases. *Plant Physiol.* **101**, 7–12 (1993).
- Negre-Salvayre, A. *et al.* A role for uncoupling protein-2 as a regulator of mitochondrial hydrogen peroxide generation. *FASEB J.* **11**, 809–15 (1997).
- Brand, M. D. in *Bioenergetics—A Practical Approach* (eds Brown, G. C. & Cooper, C. E.) 39–62 (IRL, Oxford, 1995).
- Rolfe, D. F. S., Hulbert, A. J. & Brand, M. D. Characteristics of mitochondrial proton leak and control of oxidative phosphorylation in the major oxygen-consuming tissues of the rat. *Biochim. Biophys. Acta* **1188**, 405–16 (1994).
- Miyazaki, J. *et al.* Establishment of a pancreatic beta cell line that retains glucose-inducible insulin secretion: special reference to expression of glucose transporter isoforms. *Endocrinology* **127**, 126–132 (1990).

### Acknowledgements

We thank A. Abuin for help in constructing the UCP3 knockout mice.

Correspondence and requests for materials should be addressed to M.D.B. (e-mail: martin.brand@mrc-dunn.cam.ac.uk).



## Contacts

**Publisher:** Fabien Savenay  
**Editor:** Paul Smaglik  
**Sales Director:** Ben Crowe

## European Head Office, London

The Macmillan Building  
4 Crinan Street  
London N1 9XW, UK  
Tel +44 (0) 20 7843 4961  
Fax +44 (0) 20 7843 4996  
e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

## Group European Manager:

Leonie Welss (4954)

## European Manager:

Nevin Bayoumi (4978)

## UK/ RoW/ Ireland:

Matt Powell (4953), Ben Corp (4974),  
Andy Douglas (4975)

## Holland/ Italy:

Nevin Bayoumi (4978)

## Scandinavia:

Sille Opstrup (4994)

## Spain/ Portugal:

Leonie Welss (4954)

## Production Manager:

Billie Franklin

To send films and materials use

London address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

## International

## Advertising Coordinator:

Laura Pearson (4977)

## Naturejobs web development:

Tom Hancock

## Naturejobs online production:

Ben Lund

## European Satellite Offices

## France/ Belgium:

Christine Niox-Chateau  
Tel + 33 (0) 1 43 20 16 51  
Fax + 33 (0) 1 43 20 51 52  
e-mail: [c.nioxchateau@nature.com](mailto:c.nioxchateau@nature.com)

## Germany/ Austria/ Switzerland:

Patrick Phelan/ Kate Turner  
Tel + 49 89 54 90 57 11/2  
Fax + 49 89 54 90 57 20  
e-mails: [p.phelan@nature.com](mailto:p.phelan@nature.com)  
[k.turner@nature.com](mailto:k.turner@nature.com)

## US Head Office, New York

345 Park Avenue South,  
10th Floor, New York, NY 10010-1707  
Tel +1 800 989 7718  
Fax +1 800 989 7103  
e-mail: [naturejobs@natureny.com](mailto:naturejobs@natureny.com)

## US Sales Director:

Ben Crowe

## US Sales Manager:

Peyton Mason

## US Advertising Coordinator:

Ashly de Leon

## Japan Head Office, Tokyo

MG Ichigaya Building (5F),  
19-1 Harakatomachi,  
Shinjuku-ku,  
Tokyo 162-0841  
Tel +81 3 3267 8751  
Fax +81 3 3267 8746  
e-mail: [k.cowan@naturejp.com](mailto:k.cowan@naturejp.com)  
**Japan Manager:** Kate Cowan

# naturejobs

## New Year evolutions

**P**redicting the future is never easy, but the past can often offer some clues to the way events will unfold. Several employment trends made their mark in 2001, and it is reasonable to assume that these will continue to shape the labour market over the coming year. With many fields moving from reductionist to integrationist approaches, the key for 2002 is probably a multidisciplinary approach.

As if to emphasize this trend, vast amounts of money are being poured into multi-site projects such as structural genomics. Scientists in overlapping fields would be wise to latch on to related areas or to be smart in carving out their own niche.

For several years, bioinformatics led the rising demand for mathematicians and computer scientists. The need for such skills will continue to grow as other areas — such as systems biology, ecology and geosciences — make increasing use of computer modelling. And across the pharmaceuticals and biotech sectors, there is evidence of a major shortage of biochemists.

The start of the twenty-first century has also been characterized by increased global mobility. Scientists are attracted to enabling facilities and repelled when those facilities are down. The temporary reduction in staff at CERN, the European laboratory for particle physics near Geneva, has already created an exodus to Fermilab in Illinois, which is likely to continue. And Japan's temporary closure of the Super Kamiokande neutrino experiment will almost certainly send researchers scrambling to find other facilities.

In addition, there are indications that some European countries face a shortage of candidates to replace ageing academics. If that bears out, these countries will have to fight to attract scientists who left Europe when opportunities were less ample.

**Paul Smaglik**  
*Naturejobs editor*



## [WWW.NATUREJOBS.COM](http://WWW.NATUREJOBS.COM)

Career centre  
Information on the scientific  
job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS